In silico characterization of substrate and inhibitor specificity of histone deacetylases

Colaço, Ana Rita Freitas

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Ana Rita Colaço

In silico characterization of substrate and inhibitor specificity of histone deacetylases

Combination with biochemical profilling

Supervisor: Günther H. J. Peters

Co-supervisors: Christian A. Olsen and Jens Ø. Duus

Ph.D Thesis, December 2017

DTU Chemistry
Department of Chemistry

Department: Department of Chemistry

Author: Ana Rita Colaço

Title: *In silico* characterization of substrate and inhibitor specificity of histone deacetylases: combination with biochemical profilling

Supervisor: Günther H. J. Peters

Co-supervisors: Christian A. Olsen and Jens Ø. Duus

Date: December 2017

Preface

This thesis is the result of 3 years of research, as part of the Danish program to obtain a Ph.D degree. The work has been carried out at Technical University of Denmark (DTU) under the supervision of former Associate Professor Peter Fristrup (Senior Specialist at Ørsted) and Professor Christian Adam Olsen (University of Copenhagen), and later on under the supervision of Associate Professor Günther H. J. Peters, Professor Jens Øllgaard Duus (DTU Chemistry) and Professor Christian Adam Olsen. This Ph.D scholarship was sponsored by Lundbeck Foundation and DTU.

Chapter one is a general introduction to the background of the project, including epigenetics concepts and involvement in disease, current therapies and inclusion of HDAC inhibitors as efficient alternatives. The impact and usefulness of *in silico* methods in modern drug discovery is also addressed in this chapter. Chapter two describes the in silico study developed to assess new class I HDAC inhibitors binding mode and affinity. Chapter three presents and discusses the study on sirtuin sensitivity to NAD+ and NADH. Chapter four reports the efforts made on the creation of the first SIRT7 homology model and chapter five gives a general conclusion. The experimental protocols are presented in Appendix A, supplementary tables and figures are depicted in Appendix B, and the publications in this dissertation are included as Appendix C.

Kongens Lyngby, December 2017

Ana Rita Colaço

Abstract

The field of epigenetics has received considerable attention in recent years due to its impact on genetics, developmental biology, cancer biology, and medicinal chemistry. This research project focused on histone deacetylases (HDAC), a family of epigenetic regulator enzymes that contribute to the DNA expression regulation, through hydrolysis of the ε-*N*-acylated lysine residues of histone and non-histone protein side chains.

The first chapter reports the *in silico* study performed to evaluate the binding mode and affinity of a collection of known macrocyclic HDAC inhibitors and their analogous, towards class I HDACs. To the present date, a few HDAC inhibitors have been approved for treatment of various cancers. One particularly interesting class of HDAC inhibitors is the macrocyclic peptides and depsipeptides, which are highly potent and moderate selective, and can be found in nature. My results confirmed the higher potency of hydroxamate analogues in comparison with their norvaline counterparts, as well as the crucial role of an aspartate in achieving the optimal binding position. Furthermore, the unexpected interaction between aromatic side chains of the inhibitors and the catalytic zinc ion opened a new line of investigation for possible HDAC inhibitors.

Chapters 3 and 4 focuses on class III HDACs, also known as sirtuins, a class of NAD^+-dependent enzymes with homology to the silent information regulator 2 yeast enzymes. There are seven proteins in the sirtuin family and they all share a conserved 270 amino acid catalytic domain, with variable N- and C- termini. These enzymes have been suggested as therapeutic targets for diabetes, cancer, neurodegenerative diseases and inflammation, which has led to the investigation of regulatory molecules. In Chapter 3, the NADH inhibition of SIRT1, SIRT3 and SIRT5 was investigated using long Molecular Dynamics simulations. The results showed that the addition of a single proton in the nicotinamide ring can induce an important conformational change on NADH, causing misalignment of the nicotinamide amide to the key residues in the C pocket and ultimately resulting in a significant loss of binding affinity. Chapter 4 reports the generation of the first homology model of SIRT7. With reasonable (39%) to high (100%) identity templates, this model includes the secondary structure of both N- and C- termini. The catalytic domain, similar to SIRT6, depicts a Rossmann-fold and a Zn^{2+}-binding domains. The termini show a well-defined a-helix organization, essential for the DNA binding. The model is further supported by phylogenetic and structural analysis, and secondary structure prediction.

Dansk resumé

Epigenetik har været et vigtigt omdrejningspunkt for forskning gennem de forløbende år på grund af indflydelsen på genetik, udviklingsbiologi, kræftbiologi og medicinalkemi. Dette forskningsprojekt fokuserer på histon deacetylaser (HDAC), en familie af epigenetiske regulatoriske enzymer som bidrager til DNA ekspressionsregulering, gennem hydrolyse af ε-*N*-acetylerede lysin positioner i histon samt andre protein sidekæder.

Det andet kapitel omhandler *in silico* studier udført for at evaluere bindingsmetoden og affiniteten af en samling af kendte makrocykliske HDAC inhibitorer og deres analoger til klasse I HDACer. Til dags dato er nogle få HDAC inhibitorer er blevet godkendt til behandling af forskellige kræfttyper. En yderst interessant klasse af HDAC inhibitorer er de makrocykliske peptider og depsipeptider, som er yderst potente og moderat selektive, og findes i naturen. Mine resultater bekræfter øget effekt af hydroxamat analoger i sammenligning med deres norvalin ækvivalenter, og dertil en aspartatsidekædes vigtige rolle i at opnå optimal bindingsposition. Herudover har den uventede interaktion mellem de aromatiske sidekæder fra inhibitorerne og den katalytiske zink-ion banet vejen for ny udforskning af mulige HDAC inhibitorer.

Kapitel 3 og 4 fokuserer på klasse III HDACer, også kaldet sirtuiner, en klasse af NAD+ afhængige enzymer med homologi til silent information regulator 2 enzymer fra gær.Der er syv proteiner i sirtuinfamilien, og de deler alle et konserveret 270 aminosyre katalytisk domæne, med variable N- og C- termini. Disse enzymer er blevet forslået som terapeutiske target mod diabetes, cancer, nervedegenererende sygdomme og betændelse, hvilket har ledt til undersøgelse af regulerende molekyler. I kapitel 3 blev NADH inhiberingen af SIRT1, SIRT3 og SIRT5 undersøgt ved hjælp af molekyledynamiksimuleringer. Resultaterne viste at tilføjelse af et proton til nikotinamidringen kan føre til vigtige konformationsændringer i NADH, hvilket fører til fejlplacering af nikotinamidet i forhold til nøglepositionerne i C-lommen og endeligt resultere i signifikant tab af bindingsaffinitet. Kapitel 4 rapporterer genereringen af den første homologimodel af SIRT7. Med rimelige (39 %) til høj (100 %) skabelon identitet, inkluderer denne model den sekundære struktur af både N- og C- termini. Det katalytiske domæne, ligesom i SIRT6, viser en Rossmann-fold og et Zn^{2+}-bindingsdomæne. Terminalerne viser veldefineret a-helix organisering, hvilket er essentielt for DNA binding. Denne model er yderligere støttet af fylogenetisk og strukturel analyse samt sekundær struktur bestemmelse.

Acknowledgments

I would like to thank my former supervisor Peter Fristrup and co-supervisor Christian Olsen for the oportunity to carry out the work reported in this thesis. These last three years have been a life lesson, both professionally and personally. Moving away from my home country was a big and ambicious step but it provided me with the oportunity to pursuit the research field I am passionate about and for that I am thankful. Their pacience with me and enthusiasm about the project were fruitful and thought me that smaller or bigger, achievements take time and perseverance. To my new supervisors, Prof. Günther Peters and Prof. Jens Duus, thank you so much for steping in and help me finish this journey. Your trust in my capabilities was crucial and your will to help in whatever way you could were like a safety net for me. A special thank you to Prof. Christian Olsen for the constant support throughout this project.

To the former members of the Fristrup group, Niels, Daniel Lupp, Daniel Larsen, Lasse, Allan, Mathias, Rikke and Samuel, thank you for the jokes and the laughs, thank you for the coffee breaks, thank you for the deep conversations (scientific or otherwise). More than co-workers, you became my friends and made me fill welcomed at DTU. To Niels, a special thank you for all the advises and teaching, and to Daniel Larsen, my "journey buddy", thank you for everything, your friendship made all of this a lot more fun.

I would like to thank Daniel and Sandra for proof reading parts of the thesis, and Daniel for all the computational chemistry discussions and help with organic chemistry concepts a bit out of my field. I would also like to thank Jonas Mansoor for doing his very best in keeping the clusters working and all his pacience with the software. I know how much of a pain it is and I am really grateful that you have always tryied to make the best of it.

To Christine and Faranak, the first friends I made here. You were there through happy and sad times and I will forever be grateful to you, for making me fill welcomed and supported. To Geanna, you kept me sane and your support througout this last 2 years has been more important and meaningful to me than I could ever express in words. You helped me growth as a person and become a better version of myself.

To Mathias, right when I had written-off this part of my life, we managed to somehow find eachother and that has been one of the happiest moments of this journey. It was a though year but we stood together and helped eachother up. You were the voice of reason when my

temper boiled and your pacience with my early Sundays and late nights meant the world to me.

To my friends back home and abroad, Rita, Cindy, Batista e Sandra.Studies say if a friendship lasts longer than three years, it will likely last forever. We are proof of that.Thank you for all these years and for being there even when I couldn't be.

Finally, to my family, without your support and sacrifice I wouldn't be here. To my parents who have always pushed both my brother and myself to be the very best we can be and reach for the stars. To my brother who makes me miss him and his shenanigans every day. Everything I am today and everything a achieved I owe it to you. Thank you for being the greatest family I could ever wish for.

List of abbreviations

ADPR	ADP-ribose
Aoe	2-amino-8-oxo-9,10-epoxydecanoic acid
ATRX	Alpha-thalassemia X-linked intellectual disability
bp	Base pairs
CBHA	Cinnanic bishydroxamic acid
CHAPs	Cyclic hydroxamic acid-containing peptides
CPS1	Carbamoyl-phosphate synthase
CSD	Cambridge structural database
CTCL	Cutaneous T-cell lymphoma
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMTs	DNA methyltransferases
DOPE	Discrete optimised protein energy
EMT	Epithelial-mesenchymal transition
ER	Endoplasmic reticulum
FEP	Free energy perturbation
FK-228	Dipsipeptide
FMR	Fragile X mental retardation
G6PD	Glucose-6-phosphate 1-dehydrogenase
GLUD1	Glutamate dehydrogenase 1
Glut1	Glucose transporter 1
HATs	Histone acetyltransferases
HDACs	Histone deacetylases
hGR	Human glucocorticoid receptor
HMGCS2	Hydroxymethylglutaryl-CoA synthase 2
HMMs	Hidden Markov models
HMTs	Histone methyltransferases
HPLC	High pressure liquid chromatography
IDE2	Insulin degrading enzyme 2
IGF1	Insuline like growth factors
LIE	Linear interactions energy
lnRNA	Long non-coding RNA
MBPs	Methyl-CpG-binding protein
MD	Molecular dynamics
MHC	Major histocompatibility complex
MM	Molecular mechanics
MMR	Mismatch repair
MPI	Message passing interface
NAD^+	Nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate

NAM	Nicotinamide
NF-kB	Nuclear factor-kappa B
NTDs	Amino-terminal domains
PARPs	Poly-ADP-ribose polymerases
PBC	Periodic boundary conditions
PDB	Protein Data Bank
PDH	Pyruvate dehydrogenase
PGAM	Phosphoglycerate mutase
PSSM	Position-specific score matrix
PTCL	Peripheral T-cell lymphoma
PTEN	Phosphatase and tensin homolog
PTMs	Posttranslational modifications
Rb	Retinoblastoma protein
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RNA	Ribonucleic acid
RNAi	RNA interference
ROS	Reactive oxygen species
rRNA	ribosomal ribonucleic RNA
SA	Simulated annealing
SAHA	Suberoylanilide hydroxamic acid
SAR	Structure-activity relationship
SAXS	Small-angle X-ray scatterring
SBDD	Structure-based drug design
Sir2	Silent information regulator 2
SirReals	Sirtuin rearranging ligands
SOD1	Superoxide dismutase
TCA	Tricarboxylic acid
TEA	Tetraethylammonium
TFIIE	Transcription factor II E
TI	Thermodynamic integration
tRNA	transfer RNA
TSA	Trichostatin A
ZBG	Zinc binding group

Publications

Included in Appendix

1. “Chemical Editing of Macrocyclic Natural Products and Kinetic Profiling Reveal Slow, Tight-Binding Histone Deacetylase Inhibitors with Picomolar Affinities”, Betül Kitir, Alex R. Maolanon, Ragnhild G. Ohm, Ana R. Colaço, Peter Fristrup, Andreas S. Madsen, and Christian A. Olsen, *Biochemistry* **2017**, *56*, 5134-5146.

2. “Investigating the Sensitivity of NAD^+-dependent Sirtuin Deacylation Activities to NADH”, Andreas S. Madsen, Christian Andersen, Mohammad Daoud, Kristin A. Anderson, Jonas S. Laursen, Saswati Chakladar, Frank K. Huynh, Ana R. Colaço, Donald S. Backos, Peter Fristrup, Matthew D. Hirschey, and Christian A. Olsen, *J. Biol. Chem.* **2016**, *291*, 7128-41.

Table of contents

1. Introduction

1.1 Epigenetics and disease

1.1.1 *DNA organization*

Epigenetics defines all gene expression changes that are not coded in the deoxyribonucleic acid (DNA) sequence itself.[1] DNA is the molecular basis of genetics and heredity and can be found in nearly all living cells. Its exact location depends on whether the cells possess a nucleus (eukaryotes) or lack one (prokaryotes). In humans, DNA is stored within the nucleus and is associated with histone proteins as 'beads in a string'. The formed complex of histone/DNA is called chromatin and the basic structural unit is the nucleosome. Each nucleosome is composed of ~147 base pairs (bp) of double helix DNA and an octameric core of histone proteins, with two H3-H4 dimers surrounded by two H2A-H2B dimers. A linker histone, H1, protects the internucleosomal DNA between adjacent nucleosomes.[2–4] The chromatosomes fold up to form larger-diameter fibres (~30 nm), which form loops of 300 nm length. The larger fibres are compressed and folded to produce a 250nm–wide fibre, further coiled into the chromatid of a chromosome (Figure 1).[5]

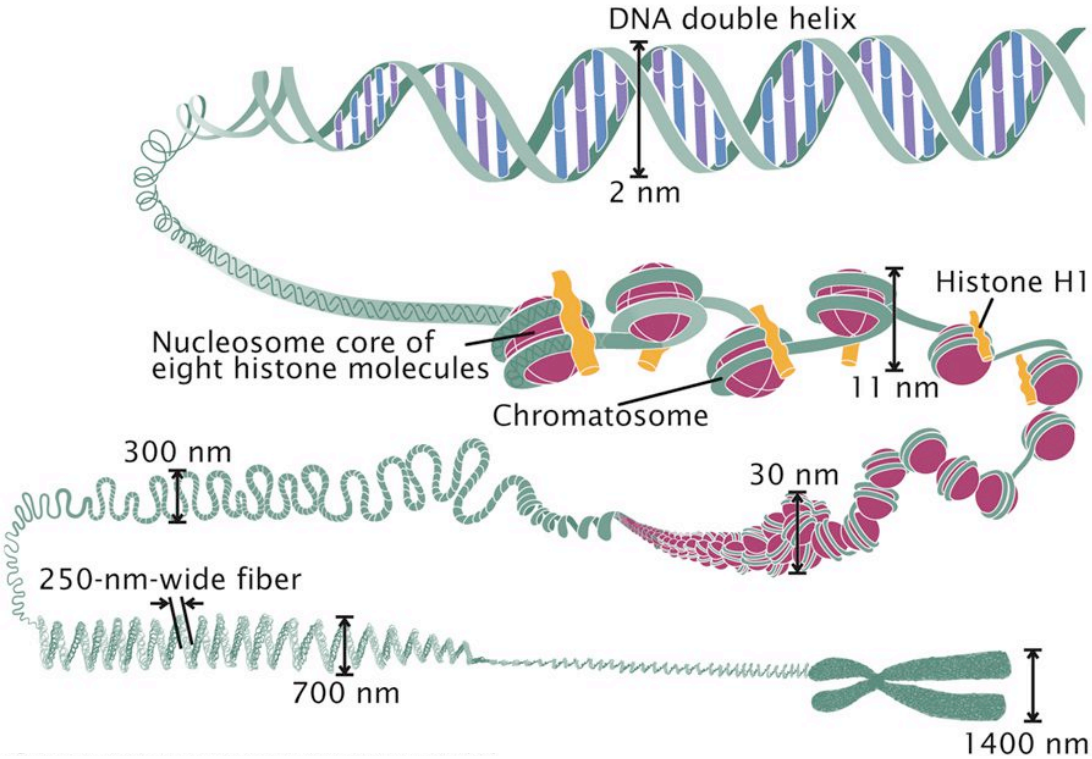


Figure 1 Chromatin organization. Adapted from [1].

Chromatin can be divided into two functional states, an open and transcriptionally active euchromatin, and a closed, tightly packaged and inaccessible heterochromatin.[6] These states affect gene expression, and thus phenotypes, and are regulated through multiple epigenetic mechanisms.[7,8] Epigenetic regulation explains cell differentiation and preservation of the phenotypes after cell division, in a multicellular organism.

1.1.2 *Epigenetic modifications*

There are three main epigenetic modifications, DNA methylation, histone modifications and RNA-associated silencing. DNA methylation inhibits gene expression by two mechanisms. The first consists of adding a methyl group to the 5-carbon position of cytosine residues on cytosine-guanine dinucleotides (CpG sites)[9,10] distributed across the human genome in regions often associated with gene regulatory regions[11]; of all CpG dinucleotides, 70-80% are methylated[12]. This highly specific modification blocks DNA recognition and binding of transcription factors, and is catalysed by DNA methyltransferases (DNMTs) 1, 3a and 3b. These enzymes function co-operatively to methylate DNA. DNMT1, the maintenance methyltransferase, recognizes hemi-methylated DNA and adds methyl groups to non-methylated daughter chromatids during replication.[10,13,14] DNMT3A and DNMT3B are called *de novo* methyltransferases and can be found in embryonic stem cells lacking DNMT1. These enzymes introduce cytosine methylation at previously unmethylated CpG sites. DNMT3L associates with DNMT3A and DNMT3B to modulate their catalytic activity. DNMT2 shows weak methyltransferase activity *in vitro*, and deletion of the corresponding gene in embryonic stem cell causes no detectable effect on the DNA methylation.[14] In the second mechanism, methyl-CpG-binding proteins (MBPs) associate with methylated cytosines blocking transcription factor access.[15–18]

Histone modifications have also been defined as epigenetic modifiers. Posttranslational modifications (PTMs) occur at the amino-terminal domains (NTDs) at the end of the eight histones in each nucleosome and include methylation, acetylation, phosphorylation, ubiquitylation, sumoylation and others (Table 1)[7]. These tails are protease sensitive and provide an exposed surface for potential interactions with other proteins.[19]

The two most well studied and best understood mechanisms are histone methylation and acetylation. Histone methylation is a chemical process that consists of adding a methyl group to lysine and arginine residues by histone methyltransferases (HMTs).[7,20] The addition of a methyl group does not change the charge of the histone tails however, the increased basicity

and hydrophobicity influences the affinity to the anionic DNA.[21,22] Methylation is associated with both transcriptionally active and inactive chromatin regions. For example, while methylation of H3K9, distributed throughout heterochromatin, marks silent DNA, methylation of H3K4 is a marker for active genes found in promoter regions.[6,23] The complexity of posttranslational histone methylation is further demonstrated by the degree of methylation. Depending if a certain amine is mono-, di- or trimethylated, there can be activation or repression of gene expression.

On the other hand, histone acetylation is usually associated with transcriptionally active chromatin and deacetylation is associated with heterochromatin.[24,25] The addition of an acetyl group to the ε-amine of conserved lysine residues by histone acetyltransferases (HATs)[20], prevents the positively charged lysine side chain from interacting with the negatively charged DNA, decreasing the affinity between histones and DNA and opening the chromatin structure. This allows transcription factors to access the promoter regions of genes. Histone deacetylases (HDACs) oppose the effects of HATs and remove the acetyl groups from lysine side chains, promoting the formation of heterochromatin and gene silencing. The balance between HATs and HDACs activities is the key to very specific acetylation levels and therefore is highly regulated.[26]

Table 1 Posttranslational modifications[7].

PTMs	**Residues Modified**	**Functions regulated**
Acetylation	**K**-ac	Transcription, Repair, Replication, Condensation
Methylation	**K**-me1, **K**-me2, **K**-me3 **R**-me1, **R**-me2a, **R**-me2s	Transcription, Repair Transcription
Phosphorylation	**S**-ph, **T**-ph	Transcription, Repair, Condensation
Ubiquitylation	**K**-ub	Transcription, Repair
Sumoylation	**K**-su	Transcription
ADP ribosylation	**E**-ar	Transcription
Deimination	**R** > Cit	Transcription
Proline isomerization	**P**-cis > **P**-trans	Transcription
Crotonylation	**K**-cr	Transcription, Repair, Replication, Condensation

Many times, histone modifications affect each other to generate specific patterns.[27,28] These observations resulted in the creation of the "histone code" hypothesis: instead of serving individual purposes, PTMs in different histone tails act in a combinatorial manner to produce a specific downstream result.[19]

Another epigenetic mechanism for transcription regulation involves gene silencing by RNA sequences. Indeed, long non-coding RNAs (lncRNAs) and RNA interference (RNAi) can lead to formation of heterochromatin and gene silencing, which is passed down through cell division. In yeast, for example, the deletion of components of the RNAi machinery impaired DNA and histone methylation of specific sites.[29–31] Although the same connection has not been shown for mammals yet, a recent study on a α-thalassaemia case reported DNA methylation and silencing of a globin gene using antisense transcription.[32]

Based on all the evidence and findings, it is clear the different epigenetic modifications are correlated and can influence each other. This allows a reliable transmission of epigenetic changes required for normal development and health, but turns the therapeutic activation of abnormally silenced genes into a much more complex and multifaceted process.

1.1.3 *Epigenetic diseases and therapies*

Disrupting any of the above mentioned mechanisms can alter the global epigenetic profiles and give rise to diseases inherited from ancestors or somatically acquired. These diseases include mental retardation, chromosomal alterations and cancer.

Fragile X syndrome is the most common inherited mental retardation and results from the expansion of the CGG repeat in the fragile X mental retardation (FMR) 1 gene untranslated region. The expanded sequence is methylated *de novo* because of the high number of CpG islands, silencing the gene and causing a 'fragile' site on the X chromosome.[33] Another example is the α-thalassemia X-linked intellectual disability (ATRX) syndrome where mutations in the ATRX gene lead to hypomethylation of ribosomal DNA, Y-specific repeats and subtelomeric repeats. Chromosomal alterations (translocations) can also lead to expression of fusion proteins that upregulate target genes. In acute promyelocytic leukaemia, an oncogenic fusion protein PML-RARα can recruit an HDAC to repress genes essential for haematopoietic cell differentiation[34].

Cancer was the first human disease to be linked to epigenetic changes, in 1983. In general, CpG islands in the promoter regions of genes are well conserved during evolution because

they are usually kept unmethylated. However, aberrant *de novo* methylation of CpG islands can occur and cause repression of tumour suppressor gene expression. This hypermethylation has been shown to constitute a hallmark of human cancer, showing in early stages of carcinogenesis.[9] In one case, researchers discovered that patients with sporadic colorectal cancers showed methylation and silencing of genes encoding MLH1, one of the seven DNA mismatch repair (MMR) proteins.[35] These aberrant methylation patterns were found not only in the tumour cells but also in spermatozoa, which can result in genetic predisposition to multiple cancers in the individuals affected.[36,37] DNA methylation inhibitors are able to reactivate gene expression when incorporated into the DNA, replacing cytosine nucleotides, during replication. Consequently, DNA methyltransferases are trapped on the DNA by covalently binding to the modified bases.[38,39] However, due to the covalent binding, high doses of these drugs can cause cell toxicity.[40]

More recently, there has been an increased study about the involvement of aberrant histone modifications in cancer development, in particular acetylation of lysine residues on histones H3 and H4. HDACs can influence gene expression through different mechanisms. Deacetylation of histone lysines represses gene expression by limiting the access of transcriptional machinery. Other studies indicate that HDACs form corepressor complexes with the nuclear receptor in the absence of a ligand. Furthermore, these enzymes interact directly with transcription factors such as retinoblastoma protein, nuclear factor-kappa B (NF-κB), transcription factor II E (TFIIE) and non-histone proteins like the tumour suppressor p53[41]. Overexpression of certain HDACs has been observed in several forms of human cancer: HDAC1 in gastric[42], prostate[43], colon[44] and breast[45] cancers; HDAC2 in cervical[46], gastric[47] and colorectal[48] cancers; HDAC3 and HDAC6 in colon and breast cancers[44,49]. Numerous studies in knockout mice have demonstrated that the different HDACs have distinct biological functions.[50] Experiments showed that overexpression of HDAC2 and -3 does not compensate for the absence of HDAC1 in knockout mice.[51,52] However, mice lacking HDAC2 survive until the perinatal period when they died from cardiac defects.[53] Treatment with HDAC inhibitors has been shown to promote gene expression, allowing cell differentiation, growth arrest and apoptosis by activating the intrinsic and extrinsic apoptotic signalling pathways[54–57]. Moreover, these inhibitors prevent angiogenesis and stimulate the immune cell recognition and activation by increased transcription of major histocompatibility complex (MHC) I and MHC II[58–61]. Some of these inhibitors, including suberoylanilide hydroxamic acid (SAHA), panobinostat, belinostat, valproic acid and romidepsin, have been approved by FDA for cancer treatment[62]. Because different families of

HDACs mediate different biological processes, the substrate specificity varies as well. For this reason, design of therapies and drugs that can target each enzyme specifically and selectively is of great importance.

1.2 HDACs and their inhibitors

The HDAC superfamily is vast and highly conserved. HDACs can be classified in four structurally and functionally different families: class I, class IIa and class IIb, class III and class IV. While class I, II and IV are Zn^{2+}-dependent enzymes, class III (referred to as sirtuins) requires NAD^{+} for deacetylation (Table 2).

Table 2 Classification of HDACs.

Classification	Mode of action	Enzymes	Homolog	Localization in cell
Class I	Zn^{2+}-dependent	HDAC1, 2, 3 and 8	Yeast Rpd3	Nucleus
Class IIa		HDAC4, 5, 7 and 9	Yeast Hda1	Nucleus and cytoplasm
Class IIb		HDAC6 and 10		Cytoplasm
Class IV		HDAC11	Yeast Rpd3 and Hda1	Nucleus and cytoplasm
Class III	NAD^{+}-dependent	Sirtuin1-7	Sir2	Nucleus and cytoplasm

Class I is closely related to yeast Rpd3 and includes HDAC1, 2, 3 and 8. These HDACs are expressed in many cell types, localized mainly in the nucleus and have been shown to be active against histone substrates. Class IIa enzymes (HDAC4, 5, 7 and 9) can shuttle from the nucleus to the cytoplasm and class IIb (HDAC6 and 10) is found in the cytoplasm.[63] Both class IIa and IIb are homologues to yeast Hda1 and their expression is restricted to specific tissues. Class IV HDAC11 shares characteristics of classes I and II HDACs and will not be mentioned further. HDACs of class I and II share a similar structural organization[64]: a single domain with a deep, tubular pocket (channel) that widens into the internal cavity where the enzyme's catalytic site is located. The channel is surrounded by several hydrophobic amino acids and the catalytic site houses a zinc ion, coordinated through histidine and aspartate residues.

Sirtuins, on the other hand, are *silent information regulator 2* (Sir2) homologues that share a conserved catalytic domain. Sirtuins are known to deacetylate lysine residues in the N-terminal domain of histone and non-histone proteins. There are seven human enzymes in this

family, assigned to five classes according to their molecular phylogeny. Sirt1, 2 and 3 belong to class I, Sirt4 and Sirt5 are the only human sirtuin representatives in class II and III respectively, and Sirt6 and 7 are included in class IV[65] (Figure 2).

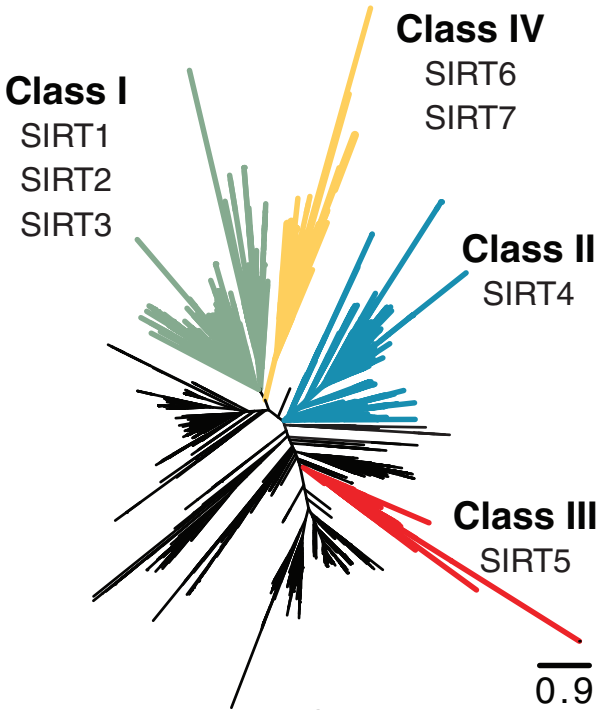


Figure 2 Phylogenetic tree of Sir2-like genes, showing the grouping of human sirtuins based on conservation patterns.[66]

1.2.1 *Class I and II HDACs*

Several small molecules have been found to inhibit class I and II HDACs. These molecules can be simpler and extracted from bacteria or fungi, or more complex and synthesised, and the IC_{50} can vary from nanomolar to micromolar concentrations. To date, a few HDAC inhibitors have been clinically approved for cancer treatment: vorinostat (SAHA **1**) for the treatment of refractory cutaneous T-cell lymphoma (CTCL)[67], romidepsin (**16**) for the treatment of CTCL and peripheral t-cell lymphoma (PTCL)[68], belinostat (**2**) for the treatment of PTCL[69] and panobinostat (**3**) as a combination therapy for patients with recurrent multiple myeloma[70]. Chiamide (**4**) is currently only approved in China as an anticancer agent. Based on their structure, HDAC inhibitors can be divided into five different classes, hydroxamates, aliphatic acids, benzamides, electrophilic ketones and cyclic peptides (Figure 3).

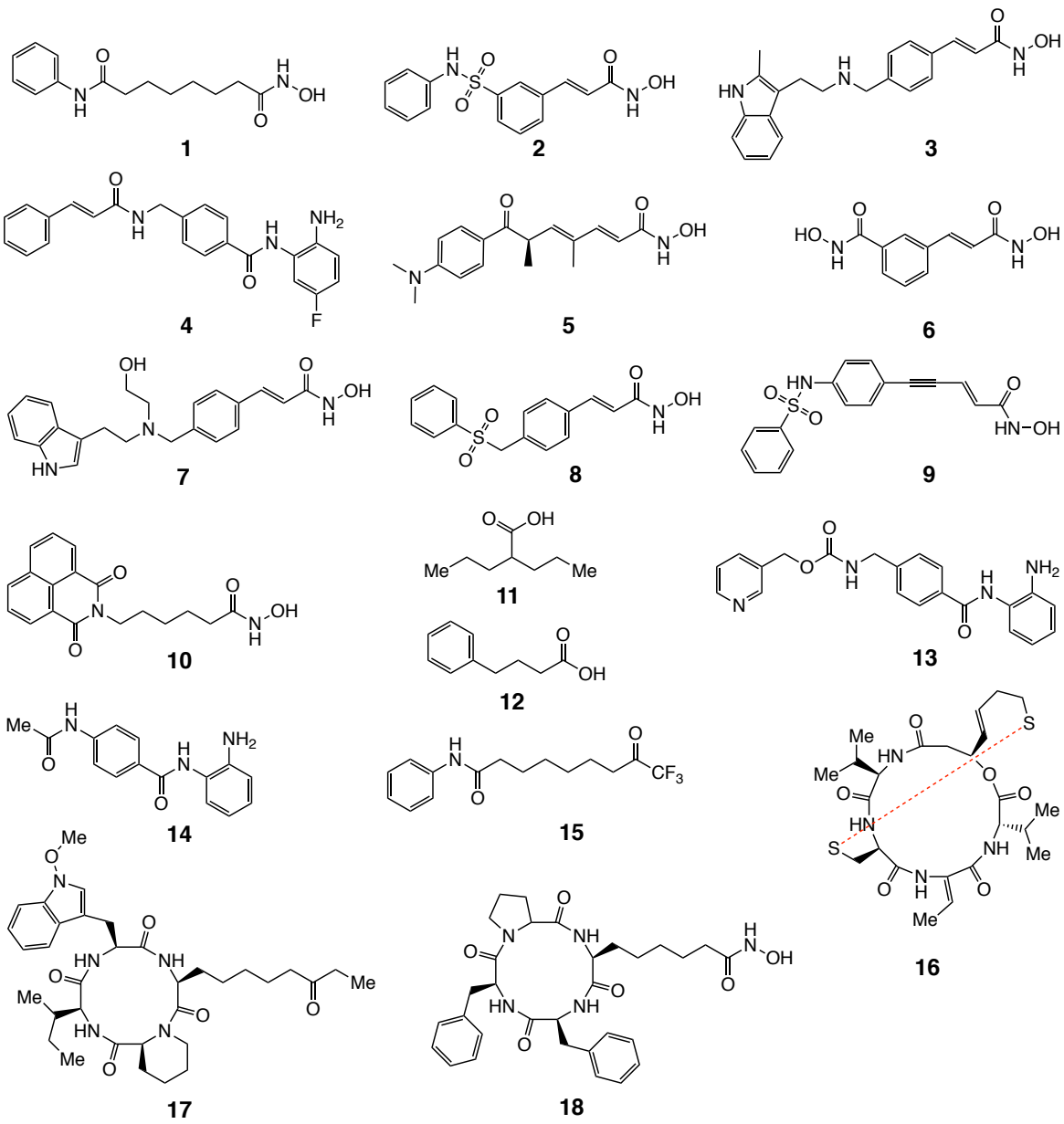


Figure 3 Class I and II HDAC inhibitors.

In 1990, trichostatin (TSA) (**5**), an hydroxamate, was the first natural product discovered to inhibit HDACs[71]. Other inhibitors in this class include SAHA (**1**), cinnanic bishydroxamic acid (CBHA)[72] (**6**) and CBHA derivatives LAQ-824 (**7**) and sulfonamide hydroxamic acid (**8**)[73,74], oxamflatin (**9**) and scriptaid (**10**)[75,76]. These inhibitors are active at nanomolar concentrations and possess longer half-lives and better bioavailability. Aliphatic acid inhibitors are the least potent of the five classes and include valproic acid (**11**) and phenylbutyrate[77,78] (**12**). The benzamide class includes MS-275 (**13**) and CI-994 (**14**) and

inhibitory activity at micromolar level[79,80]. Electrophilic ketones inhibitors possess HDAC inhibitory activity in the same range as benzamides. These molecules include trifluoromethyl ketones (**15**) and α–ketoamides[81]. The cyclic peptides class is the most structurally complex and includes depsipeptide (FK-228)[82] (**16**), apicidin[83] (**17**) and the cyclic hydroxamic acid-containing peptides (CHAPs)[84] (**18**). Like the hrydroxamate class, cyclic peptides are strong inhibitors of class I and II HDACs. This class of inhibitors will be the focus of Chapter 2 in this thesis.

The study of the X-ray crystallographic structure of HDAC with TSA[64] offered an explanation for the potency of inhibitors, detailing the inhibitor-enzyme interactions. These data showed an active site with a tubular pocket of 11Å depth that constricts to ~4.5Å halfway down but opens up to an internal cavity adjacent to the tunnel. The tunnel is surrounded by hydrophobic and aromatic residues and the side chains of two phenylalanine residues, positioned facing each other, mark the most slender part of the tunnel. The zinc ion is located in the bottom of the pocket and is coordinated by two aspartates, a histidine and a water molecule, arranged in a tetrahedral geometry (Figure 4).

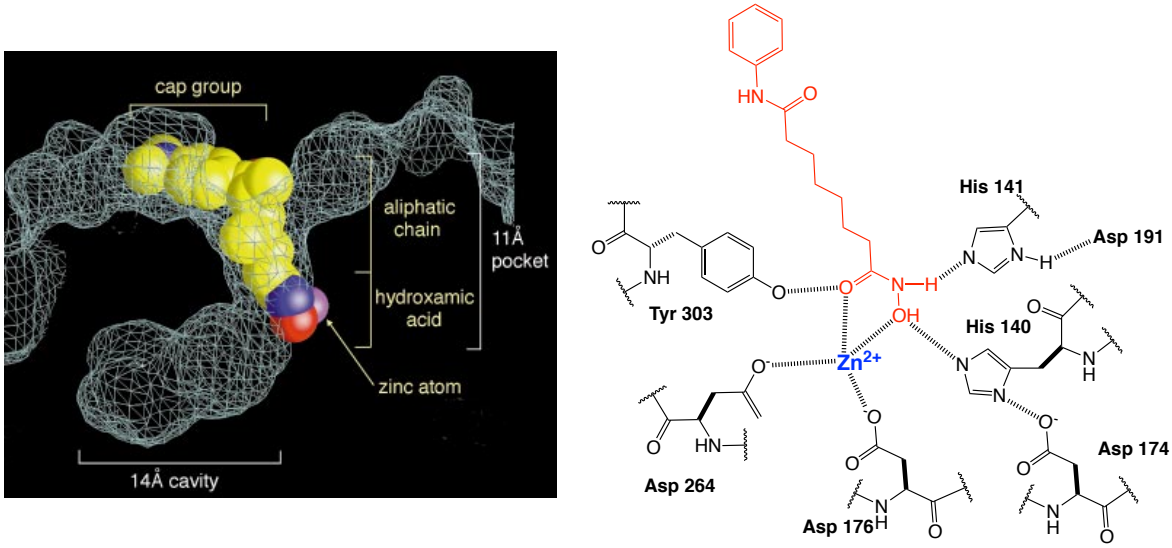


Figure 4 Representation of HDAC pocket with SAHA bound.

Regarding TSA, the hydroxamic acid coordinates the catalytic zinc, while the aliphatic chain allows numerous van der Waals interactions with the hydrophobic residues around the channel. The cap group blocks and closes the active site by interacting with the enzyme surface. This information served as basis to create a model pharmacophore to help the design of new inhibitors. The pharmacophore consists of three regions (Figure 5). The zinc chelating

group (hydroxamic acids, carboxylic acids, benzamides, epoxides), represented in red, displaces the water molecule and coordinates to zinc; the linker (in black) spans the length of the tubular pocket, interacting with the hydrophobic residues, and connects to a cap group (in dark blue) that blocks and closes the active site by interacting with the enzyme surface. This group can be a small planar group or a tetrapeptide which leads to more potent inhibitors[84,85].

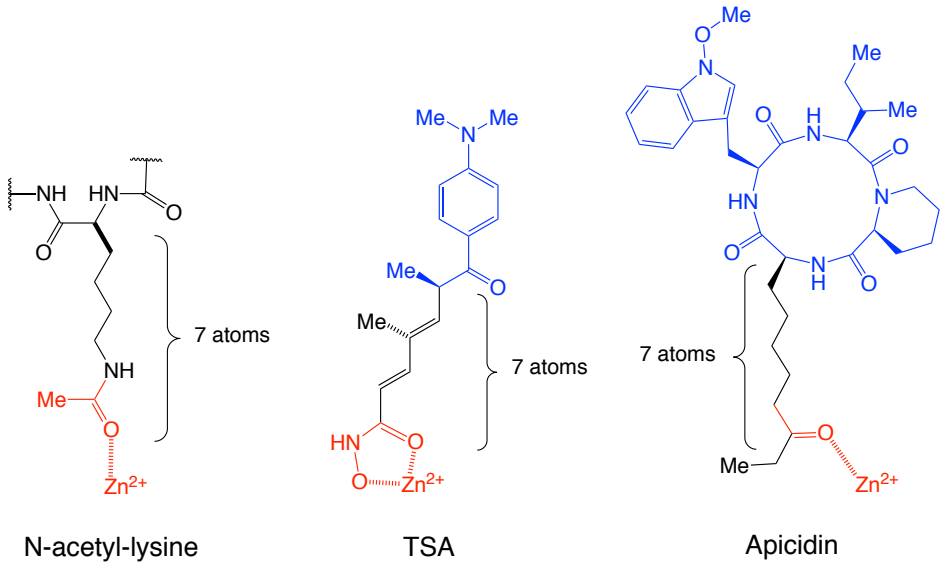


Figure 5 Pharmacophore model for HDAC inhibitors.

1.2.1.1 Cyclic peptides inhibitors

Cyclic peptides are among the most potent and structurally complex HDAC inhibitors and, in general, are known to bind in a similar manner to TSA. Structural variation have been studied within the cap group and the zinc binding group (ZBG). Based on the macrocyclic moieties, these inhibitors can be divided into two classes, cyclic tetrapeptides and bicyclic depsipeptides.

The first class includes macrocycles with both proteinogenic and non-proteinogenic amino acids of *L*- and *D*- configuration, and the 13-membered azumamides with a β-amino acid. The majority of the macrocycles of this class contain at least one aromatic side chain (except HC-toxin) and a cyclic amino acids like proline and pipecolic acid, intended to reduce the macrocycle ring strain and facilitate the ring closure[86]. Moreover, all the cyclotetrapeptides contain an amino acid with an aliphatic chain linker connected to a ZBG. The most commonly used ZBG are α-epoxyketones, α-hrydroxyketones, ketones, amides, carboxylic acids, hydroxy groups, diols and sulfenic acids.

Some of the most well-known examples of cyclic tetrapeptides include traproxin A/B[87] (**19/20**), Cyl-1/2[88] (**21/22**), chlamydocin[88–90] (**23**), HC-toxins[91,92] (**24**) and WF-3161[93,94] (**25**) (Figure 6). They all contain a 2-amino-8-oxo-9,10-epoxydecanoic acid (Aoe) amino acid with an epoxy-ketone as ZBG, which reacts irreversible with class I HDACs, with the exception of HDAC8[95]. These Aoe containing macrocycles showed a high selectivity towards class I HDAC and inhibitory activity in the nanomolar and picomolar range[84]. The replacement of the epoxy-ketone with other groups led to reduction of activity (complete loss of activity in the case of chlamydocin if the replacement is an alcohol)[96].

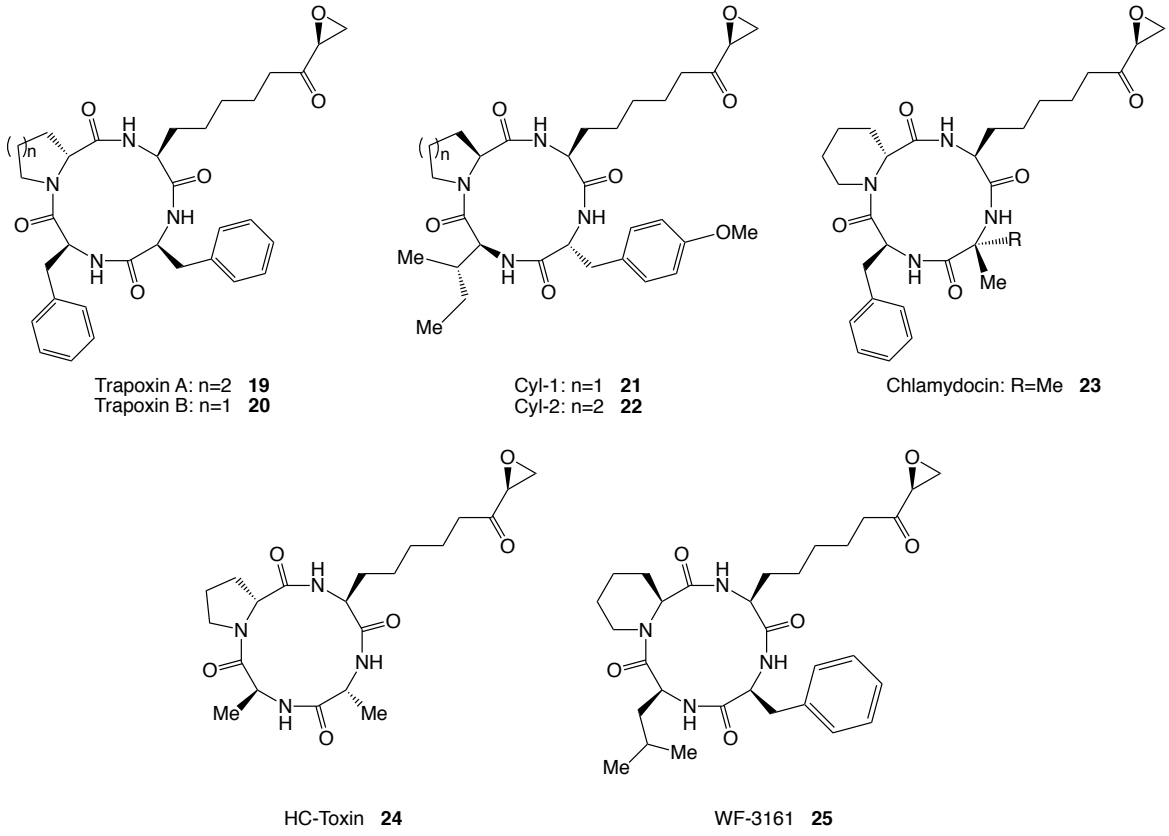
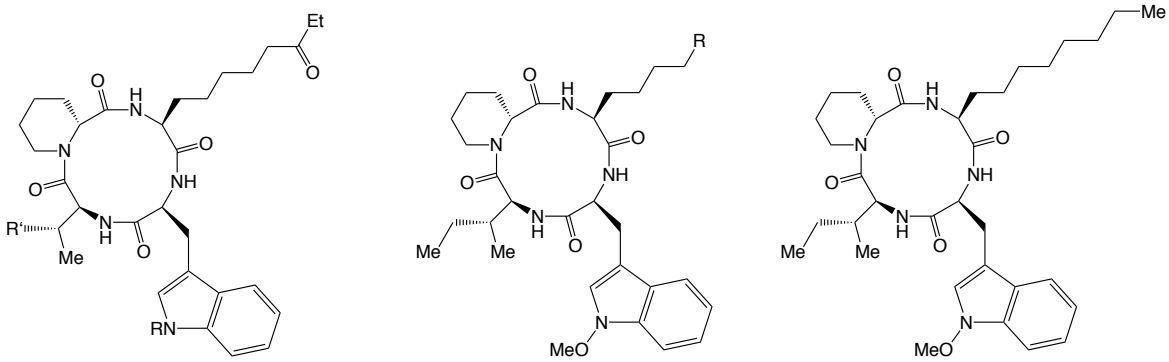


Figure 6 Natural tetrapeptide HDAC inhibitors.

Apicidin (Figure 7), another cyclic tetrapeptide, lacks the epoxy-ketone and possesses instead an ethylketone ZBG (**26-28**). Although the binding in this case is reversible, apicidin still showed selectivity to HDAC1 and nanomolar inhibitory activity. Structure-activity relationship (SAR) studies performed by Colletti and co-workers[97] demonstrated the importance of the C-8 keto group, as well as the importance of the tryptophan residues in the macrocycle. They found that reduction of the C-8 keto group to alcohol or CH2 resulted in decrease of the inhibitory activity. However, substitution of the entire ethyl ketone group

with hydroxamic acid, epimeric epoxide and hydroxyketone, improved the activity relative to apicidin. Similarly, replacement of the methoxy group in the tryptophan side chain with an isosteric ethyl moiety decreased the activity. Yet, substitution with methyl retained the activity, and removal of the methoxy group slightly increased the biological activity. Results from Singh et al.[83] confirm the previous observations, with apicidin A (**27**) and D_1 (**29**) being more active compounds (IC_{50}= 1-4nM), while apicidin D_2 (**30**) was considerably less active (IC_{50} > 400nM).



Apicidin: R=OMe, R'=CH_2CH_3 **26**
Apicidin A: R=H, R'=CH_2CH_3 **27**
Apicidin C: R=OMe, R'=CH_3 **28**

Apicidin D_1: R=$CH_2COCH(OH)CH_3$ **29**
Apicidin D_2: R=$CH_2CH(S\text{-}OH)CH_2CH_3$ **30**
Apicidin D_3: R=$CH_2CH_2CH(OH)CH_3$ **31**

Apicidin E **32**

Figure 7 Apicidin analogues.

A series of analogues was synthesised by Komatsu and co-workers[84,98], where the epoxy-ketone group was replaced by an hydroxamic acid as the ZBG in trapoxin A/B, WF-3161, chlamydocin and HC-toxin. All the hydroxamic acid compounds in the study were potent inhibitors of HDAC1.

The conformation of the macrocycle backbone proved to be of great importance for the affinity to HDACs and varies depending on the solvent. Structural data obtained for apicidin in $CDCl_3$ revealed all the bonds to be *trans*-amide, while in dimethyl sulfoxide (DMSO) the amide bond between the pipecolic acid and isoleucine residue rotated to a *cis* configuration. This conformation is also present in the X-ray crystal structure. The impact of the two different conformations on the inhibitory activity was studied by Ghadiri's group. They provided evidence that prove *cis-trans-trans-trans* to be the most potent conformation of apicidin.[99]

Azumamides are structurally similar to apicidins except for the cyclic amino acid in the macrocycle backbone. These particular inhibitors incorporate three R-amino acids and one

unique beta-amino acid in azumamide A (**33**), B (**34**) and D (**35**), and a free acid in C (**36**) and E (**37**) (Figure 8). Azumamides are potent HDAC inhibitors with an approximately 100-fold selectivity for class I over class II HDACs. Azumamide E with a hydroxamic acid zinc binding group showed similar activity to romidepsin[100].

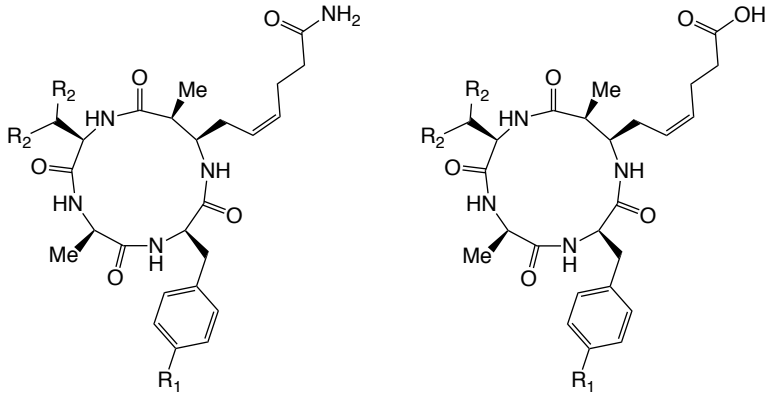


Figure 8 Structure of natural compounds azumamides.

The second class of HDAC inhibitors includes sulphur-containing bicyclic depsipeptides isolated from bacterial fermentation extracts, such as largazole (**38**) that features a thioester to form the second ring, and romidepsin (**16**), FR901375, spiruchostatins (**39-42**) and thailandepsins (**43-45**) containing a disulphide bridge instead[101] (Figure 9). Although these inhibitors do not seem to satisfy the usual phamacophore, these molecules are believed to act as prodrugs and need to go through intracellular reactions to expose the thiol group side chain and interact with the catalytic zinc[82,102]. This prodrug concept protects the thiol group from further degradation and improves membrane permeability[82,103]. Biological data showed these compounds are potent inhibitors for class I HDACs with IC_{50} values in the low nanomolar level, but only moderate inhibitors of class IIb HDACs.

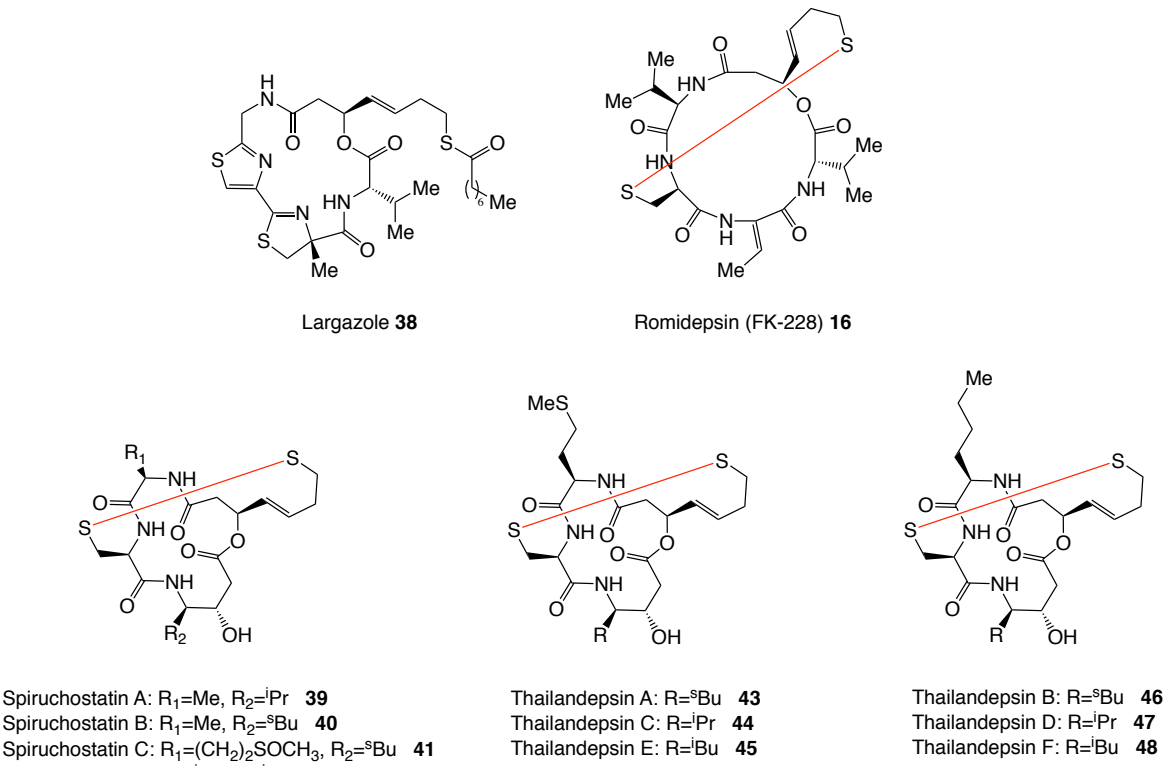


Figure 9 Sulphur-containing bicyclic depsipeptides.

Similarly to cyclic tetrapeptides, the stereochemical configuration of depsipeptides is also crucial for their inhibitory activity. Several stereoisomers of largazole have been tested for HDAC inhibition and showed weak potency compared to the parent compound. The length of the linker and the ZBG were modified and studied as well. Extension and shortening of the carbon linker reduced significantly the potency of largazole against HDACs. Incorporation of benzamides and thioacetamides as ZBG changed the selectivity for HDAC1 and HDAC6, respectively. At the same time, use of a carboxylic acid in place of the thiol suppressed all inhibitory activity.[101]

1.2.2 *Class III HDACs (Sirtuins)*

Sirtuins have been known to play crucial roles in several cellular processes such as transcription, fat mobilization, apoptosis, genomic integrity, metabolism, circadian rhythm, autophagy and life span extension[104–110]. Their clinical relevance has become widely recognized and they have been suggested as therapeutic targets for obesity, diabetes, cancer, neurodegenerative diseases and inflammation (Figure 10). This has led to the investigation of

regulatory molecules that might control their activity, which requires the understanding of the structure, chemistry and substrate specificity.

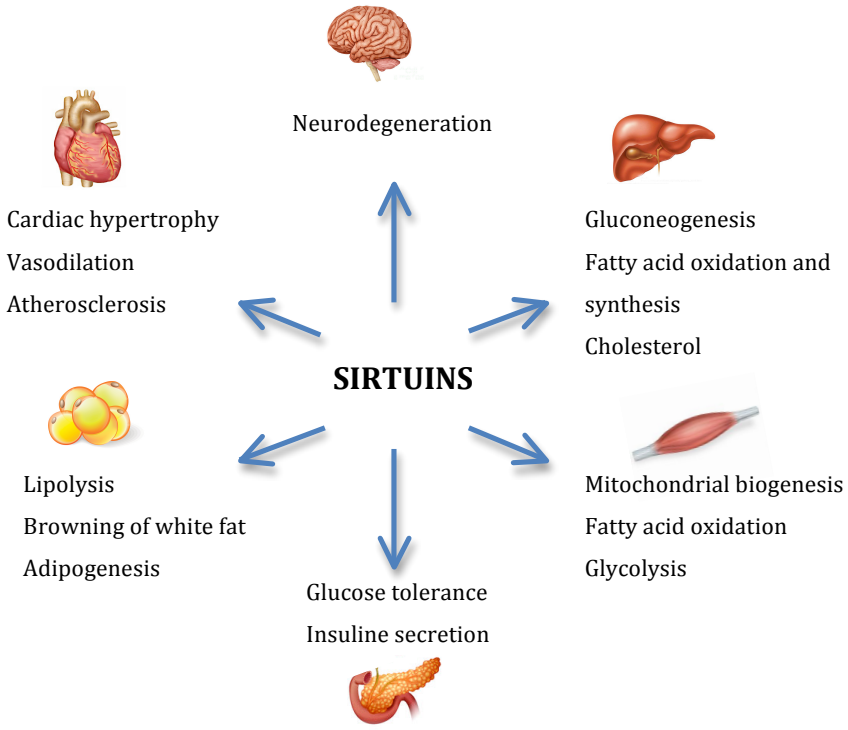


Figure 10 Sirtuins in metabolism.

1.2.2.1 Structure

As mentioned before, sirtuins share a conserved catalytic core divided in two domains (Figure 11). The larger domain constitutes a Rossmann fold, a binding motif found in enzymes that bind NAD^+/NADH or $NADP^+$/NADPH. The smaller domain is linked to the Rossmann fold by two loops, one of which the FGE loop, and includes the binding of a zinc by four highly conserved cysteine residues.[111] Between the two domains there is a structural cleft where the acylated peptide and NAD^+ bind, this brings the domains closer together and closes the cleft.[112] Interactions between the substrate main chain and both FGE loop and Rossmann fold, position the aliphatic chain of the acyllysine in the hydrophobic tunnel leading to the buried active site.[112] Similarly to other classes of HDACs, this tunnel is surrounded by conserved residues that established van der Waals interactions with the lysine aliphatic side chain. The acyl modification is also crucial for the substrate binding, studies performed with an unacylated peptide showed a 1000-fold decrease in binding affinity[113].

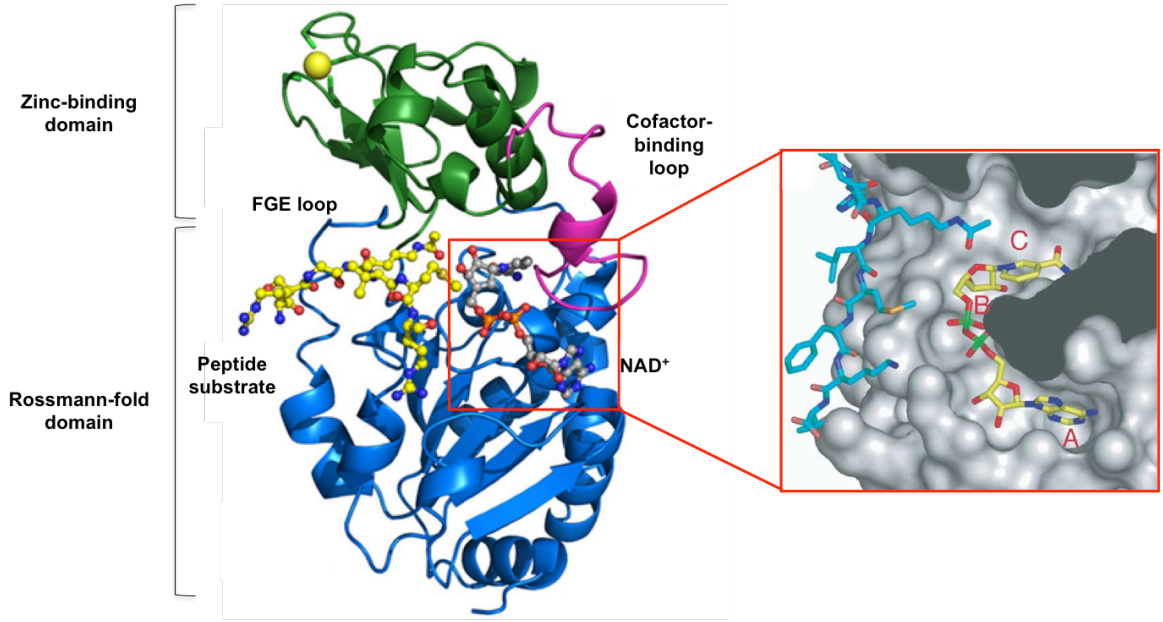


Figure 11 Sirtuin structure with zoom in on the NAD+ binding site.

NAD+ binds, in a separate pocket adjacent to the hydrophobic tunnel, right after the substrate binding induces conformational changes in the enzyme. This phenomenon promotes the binding of NAD+ in the productive conformation, with the nicotinamide buried in the C pocket[114]. The binding and positioning of NAD+ in the binding pocket will be further discussed in Chapter 3.

1.2.2.2 Deacylation mechanism

The most common reaction catalysed by sirtuins is the deacetylation of substrates. In this reaction, the NAD+ nicotinamide moiety is cleaved and the acetyl group is transferred from the lysine substrate to the ADP-ribose, forming 2'-O-acetyl-ADP-ribose. Sirtuins have also been shown to catalyse the removal of other acyl modifications, using the same main chemical reaction, as long as the enzyme is capable of accommodating the acyl groups in the binding pocket. For example, Sirt1-3 can remove propionyl and butyryl groups[115,116], while Sirt5 removes succinyl, glutaryl and malonyl groups[117]. For this reason, sirtuins should be referred to as deacylases instead of deacetylases.

In the first step of the deacylation mechanism (Figure 12), the amide bond of the acyl lysine is used as the nucleophile to attack the C1'-position of the nicotinamide ribose in NAD+. Nicotinamide is released and a C1'-*O*-alkylamidate intermediate is formed (intermediate I). A

conserved histidine residue is used as general base to deprotonate the ribose 2'-OH which then attacks the alkylamidate carbonyl carbon, forming 1'-2'-cyclic intermediate (intermediate II). This cyclic intermediate is further hydrolysed to 2'-O-acyl-ADP-ribose (2'-*O*AADPr). This step however, lacked direct experimental evidence. Earlier this year, Lin and co-workers were able to crystallise a SIRT2 complex with a thiomyristoyl inhibitor where a novel intermediate 1′-SH-2′-O-alkylimidate (III) was captured (Figure 12).[118] The data suggests that intermediate II is likely to decompose to 1′-OH-2′-O-alkylimidate intermediate (III) which can then be hydrolysed and form a tetrahedral intermediate (IV). This is further decomposed to form 2'-OAADPr and deacylated lysine.

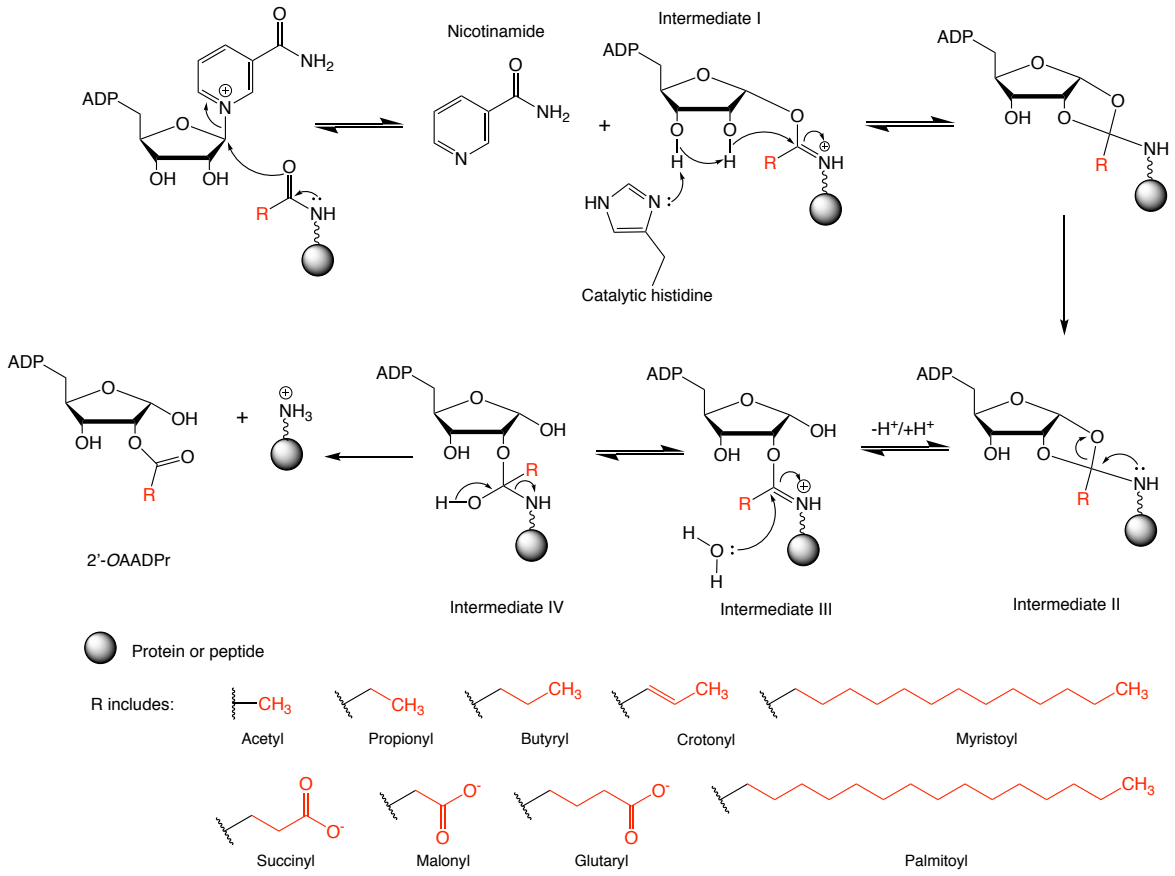


Figure 12 Sirtuin catalysed deacylation.

1.2.2.3 Substrate specificity and biological role in cancer

Although all human sirtuins share a conserved catalytic domain, differences among their sequences underlie their distinct subcellular localization, substrate specificity and biological functions (Table 3). For example, many of SIRT1, 2, 6 and 7 substrates are nuclear or cytosolic proteins, which is consistent with their known localizations. SIRT1, mostly localized

in the nucleus but also present in the cytoplasm, regulates chromatin organization by histone deacetylation, and the activity of transcription factors and other epigenetic enzymes by protein deacetylation. Among its substrates, deacetylation of H3K9, H3K14, H4K16 and H1K26 by nuclear SIRT1 promotes the formation of heterochromatin and gene silencing, and deacetylation of H3K56 helps maintaining genome stability. SIRT1 can shuttle from the nucleus to the cytoplasm and regulate tumour genesis by inhibiting tumour suppressors like p53, retinoblastoma protein (Rb) and phosphatase and tensin homolog (PTEN), and promoting cell mobility, migration and invasion (cortactin), and expression of oncoproteins such as c-Myc and Ras. SIRT2 is mainly found in the cytoplasm where it regulates cell cycle progression and tumour genesis by deacetylating α-tubulin and keratin 8. During mitosis, SIRT2 can shuttle to the nucleus and promote genome stability through H4K16 deacetylation. SIRT2 has also important roles in glucose homeostasis *via* deacetylation of glucose-6-phosphate 1-dehydrogenase (G6PD) and phosphoglycerate mutase (PGAM). SIRT6 and 7 are nuclear chromatin-bound enzymes. SIRT6 regulates inflammation *via* deacetylation of H3K9 at the NF-κB target promoters, and inhibits ribosome biogenesis by deacetylating H3K56 and repressing MYC transcription. SIRT6 also regulates glucose homeostasis by inhibiting glycolytic enzymes like the glucose transporter glucose transporter 1 (Glut1). These findings support the tumour suppressor activity of SIRT6. Besides deacetylation, SIRT6 can also remove myristoyl groups from TNFα K19 and K20, increasing its secretion. Overexpression of SIRT6 was also shown to extend mice lifespan through regulation of insulin like growth factor 1 (IGF1) serum levels. SIRT7 catalyses selective deacetylation of H3K18, an emerging epigenetic biomarker of aggressive tumours and poor clinical outcome in patients with cancer. SIRT7 intervenes in the ribosomal ribonucleic RNA (rRNA) and transfer RNA (tRNA) synthesis, promoting ribosome biogenesis necessary for tumour cell growth and proliferation. Finally, SIRT7 is reported to inhibit signalling by the tumour suppressor p53.

On the other hand, most of SIRT3, 4 and 5 substrates are mitochondrial proteins, consistent with their mitochondrial localization. SIRT3 is the strongest mitochondrial deacetylase and increases fatty acid oxidation, mitochondria biogenesis and ATP synthesis by deacetylating metabolic enzymes in these pathways. SIRT3 can also repress the production of reactive oxygen species (ROS). SIRT4 is tought to act as a lipoamidase[119], ADP-ribosyl transferase[120] and deacetylase[121]. This enzyme can remove lipoyl groups from enzymes that catalyse oxidative decarboxylation of α-keto acids such as pyruvate dehydrogenase (PDH). SIRT4 can negatively regulate mitochondrial glutamine metabolism by mono ADP-ribosylation of glutamate dehydrogenase 1 (GLUD1), which results in blocking glutamine metabolism and

tricarboxylic acid (TCA) cycle.[120] Moreover, SIRT4 can also deacetylate malonyl-CoA decarboxylase at K471, regulating lipid homeostasis.[121] A recent study by Hirschey and co-workers has shown that SIRT4 is able to remove methylglutaryl-, hydroxymethylglutaryl- and 3-methylglutaryl-lysine modifications from enzymes in the leucine metabolism pathway, restoring the catabolism of leucine.[66] SIRT5 preferentially targets negatively charged lysine modifications, including malonylation, succinylation and glutarylation. SIRT5 desuccinylates carbamyl-phosphate synthase (CPS1), superoxide dismutase (SOD1) and hydroxymethylglutaryl-CoA synthase 2 (HMGCS2) activating them and regulating blood ammonia levels, decreasing production of ROS, and modelling ketogenesis, respectively. Furthermore, SIRT5 can positively regulate glycolysis through demalonylation of glycolytic enzymes.

Table 3 Summary of some sirtuin targets referred in this chapter.

Sirtuin	Localization	Substrates/Targets	Function
SIRT1	Nucleus	H3K9, H3K14, H3K56, H4K16, H1K26	Chromation regulation and transcription
		p53, c-Myc, Rb	DNA repair and cell survival
		FOXO1, FOXO3a, PPARα	Metabolism
SIRT2	Cytoplasm	α-tubulin, keratin 8	Differentiation
		G6PD	Metabolism
	Nucleus	H4K16, H3K56, H3K18	Cell cycle
SIRT3	Mitochondria	LCAD, VLCAD, SOD2, ACECS2	Metabolism
SIRT4	Mitochondria	GDH, PDH, IDE, MCD	Metabolism
SIRT5	Mitochondria	CPS1, HMGCS2, SOD1	Metabolism
SIRT6	Nucleus	H3K9, FOXO3	Chromatin and DNA repair
		TNFα, c-Myc, HIF1α	Metabolism
SIRT7	Nucleus	H3K18, p53, RNA Pol 1	Transcription

Due to the abundance of the cellular partners and the complexity of the metabolic network in which these enzymes are involved, sirtuins can affect the tumour microenvironment. By regulating multiple aspects like inflammation, hypoxia, epithelial-mesenchymal transition

(EMT) and angiogenesis, sirtuins can have tumour-promoting or tumour-suppressive roles (Table 4). One example is SIRT1. Previous studies showed SIRT1 is downregulated in various human cancers like breast and ovarian cancers, liver cancer, prostate cancer and bladder carcinoma. In the case of breast and ovarian cancers with BRCA1 mutations, this transcription factor can no longer bind SIRT1 and inhibit the expression of surviving. The result is loss of SIRT1 positive regulation and increased expression of surviving with formation of mammary tumours.[122] On the other hand, SIRT1 can also be oncogenic and display elevated levels in colon cancer and haematopoietic cancers. Overexpression of SIRT1[123–126] has been shown to promote the development of thyroid and prostate cancers in Pten+/- mice, by promoting expression of c-Myc[127].

Table 4 Sirtuins in human cancers.

Sirtuin	Cancer type	Sirtuin expression	Sirtuin role	Refs
SIRT1	Breast, liver, glioblastoma, bladder, prostate, ovarian, colon, skin	Reduced	Tumour suppressor	128–130
	Colon, thyroid, acute/chronic myeloid leukaemia, liver	Increased	Oncogene	123–126,131,132
SIRT2	Breast, liver, glioblastoma, prostate	Reduced	Tumour suppressor	133
SIRT3	Breast, ovarian, lung, testicular glioblastoma, prostate, liver pancreatic	Reduced/Deleted	Tumour suppressor	134–137
	Brain, oral squamous cell	Increased	Oncogene	138,139
SIRT4	Gastric, bladder, breast, leukaemia, lung	Reduced	Tumour suppressor	140
SIRT5	Non-small-cell lung	Increased	Oncogene	141
SIRT6	Pancreatic, colon, liver	Reduced	Tumour suppressor	142–144
	Breast	Increased	Oncogene	145
SIRT7	Uterine, colon, kidney, ovarian, prostate	Increased	Oncogene	146,147

Studies like the previously mentioned help discover specific conditions in which a given sirtuin can be activated or inhibited, with small molecules, and contribute to combat cancer progression.

1.2.2.4 Sirtuin inhibitors

The development of small molecules that can influence sirtuin activity has been the object of increasing interest and during the last decade, both activators and inhibitors have been developed. Here, we will focus on the inhibitors.

Nicotinamide (**49**) is an endogenous inhibitor of all human surtuins, with IC_{50} values varying from 1 to 100 uM.[148,149] Nicotinamide can bind on the sirtuin C pocket and reverse the deacylation reaction by nucleophilic attack of the pyridine nitrogen on the O-alkylimidate intermediate. Nicotinamide has been shown to block proliferation and promote apoptosis in leukemic cells, and inhibit growth and viability of human prostate cancer cells[150,151]. Modifications on the nicotinamide led to the development of several 2-anilinobenzamide analogues (**50, 51**). Analogue 8 is SIRT1 selective[152] while analogue 9 selectively targets SIRT2, inducing increase of α-tubulin acetylation in human colon cancer cells (Figure 13).[153] Dihydropyridines (**52**) can also inhibit sirtuins and SAR studies showed substitution at the N1 position with cyclopropyl, phenyl and phenylethyl targets SIRT1 and 2, but not SIRT3 (Figure 13).[154]

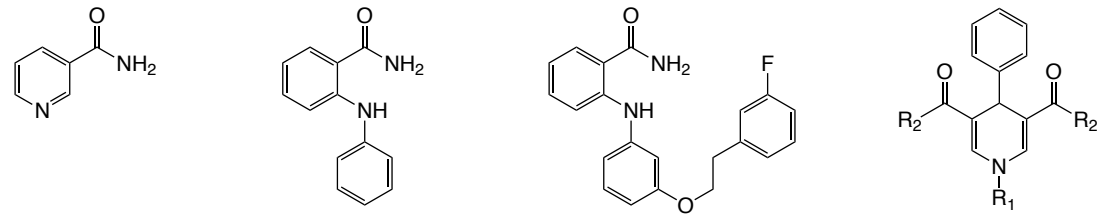
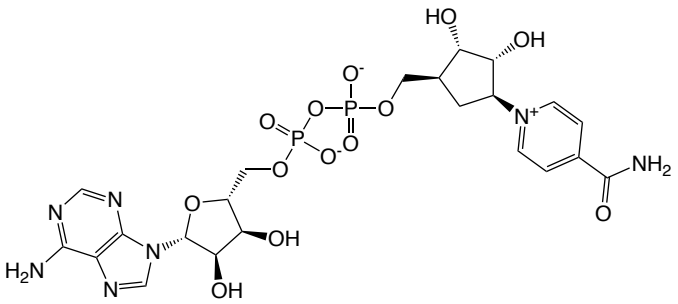


Figure 13 Structure of nicotinamide and analogues, and dihydropyridine.

Carba-NAD^+ (**53**) (Figure 14) was one of the first sirtuin inhibitors to be identified. This compound is thought to bind in the same manner as NAD^+ but with the replacement of the nicotinamide-ribose oxygen with a carbon atom, the C-1' loses electrophilicity and the nicotinamide exchange reaction is blocked.[155] However, the low cell permeability makes

carba-NAD+ not suitable for cellular studies while its high water solubility is excellent for co-crystallization with sirtuins.[156,157]



Carba-NAD+ **53**

Figure 14 Structure of sirtuin inhibitor Carba-NAD+.

Another class of hydroxynaphthaldehyde derivative sirtuin inhibitors, includes sirtinol (**54**), salermide (**55**), cambinol (**56**) and respective analogues (**57-59**) (Figure 15). Most of these compounds are SIRT1 and SIRT2 inhibitors with affinities at the uM level.[158] Sirtinol (**54**) does not change the levels of global acetylation of histones and tubulin in cells but multiple studies showed anticancer activity for this compound. Sirtinol (**54**) can induce senescence-like growth arrest through activation of RAS-MAPK pathway and cell apoptosis through p53 action[159]. Salermide (**55**) is a SIRT2 selective inhibitor and induces apoptosis in several human cancer cell lines such as, leukaemia cell lines, colon cancer, lymphoma and breast cell lines.[160] Cambinol (**56**) is a sirtuin inhibitor with similar potency towards SIRT1 and SIRT2 and weak inhibition against SIRT5. This compound competes with the lysine substrate and the beta-naphthol seems to be essential for its activity. Substitutions on the phenyl ring or functionalization of N1-position improve activity and selectivty: *p*-bromo analogue increases selectivity for SIRT1, while N-alkylation with butyl (**57**) improves selectivity for SIRT2[161]. A pyrazolone-based cambinol analog 8 (**58**) showed selectivity towards SIRT3[162]. Cell based studies showed cambinol also promotes hyperacetylation of α-tubulin, p53, KU70 and FOXO3a. Thiobarbiturates, derived from the thiouracil of cambinol, were discovered as a novel class of sirtuin inhibitors. One of the members of this class, the arylidene-thiobarbiturate (**59**), is a potent inhibitor of SIRT5 desuccinylation[163].

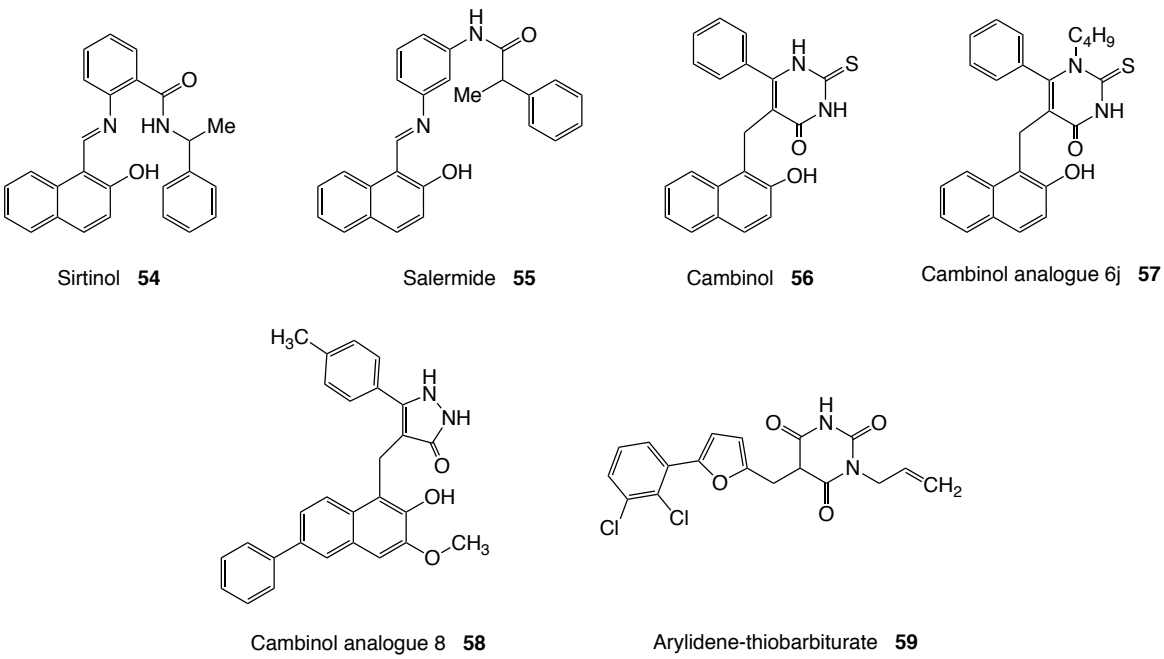


Figure 15 Structure of hydroxynaphthaldehyde derivative inhibitors.

EX-527 (**60**) (Figure 16), an indole, was first SIRT1 inhibitor to be discovered and, up until today, the only one that reached clinical trials. This compound depicts a carboxamide moiety that mimics the nicotinamide amide group. Kinetic studies showed Ex-527 (**60**) acts as a uncompetitive inhibitor to NAD^+ and non-competitive to the substrate[164]. EX-527 (**60**) has been tested in cellular and animal models and revealed to be the most promising in the treatment of Huntigton's disease.[165] Several analogues based on EX-527 (**60**) have been synthesized but did not show improved potency or selectivity.

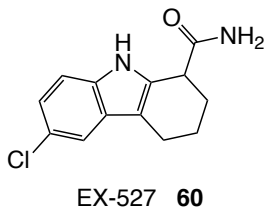
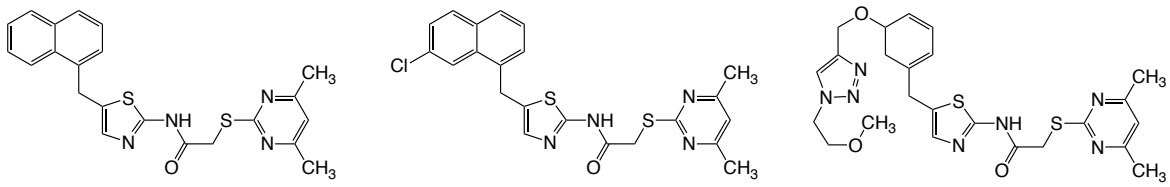


Figure 16 Structure of EX-527 inhibitor.

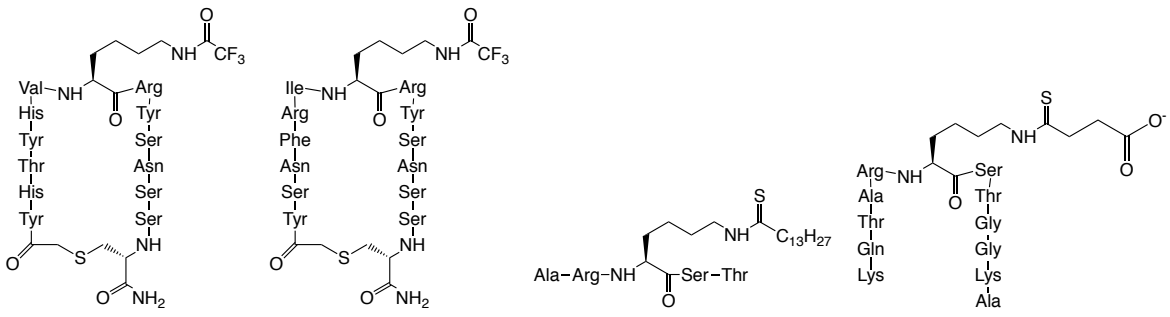
Sirtuin rearranging ligands (SirReals) (**61-63**) were discovered through library screening and co-crystallization with SIRT2 showed a rearrangement of the active site upon ligand binding. These inhibitors proved to affect SIRT2 catalysed reactions[166] and promote tubulin acetylation on cancer cell lines[167]. The structures of these inhibitors are depicted in Figure 17.



SirReal2 **61** SirReal analogue 14a **62** SirReal analogue 10 **63**

Figure 17 SirReal sirtuin inhibitors.

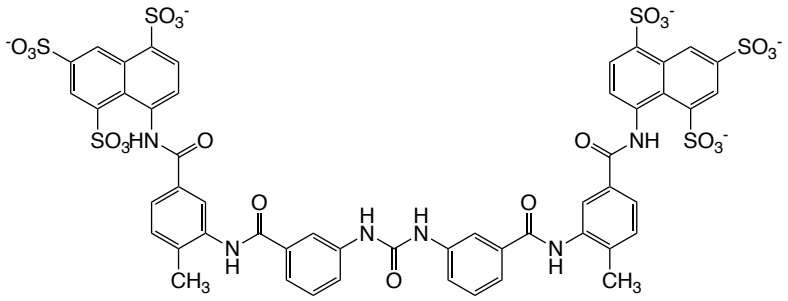
Thioacyllysine-containing compounds (**64-67**) (Figure 18) mimic the sirtuin substrate and bond covalently to ADP-ribose during the first step of the sirtuin deacylation reaction. This intermediate is more stable and therefore resistant towards deacylation, blocking the enzymatic activity.[168,169] This kind of inhibitor has been thoroughly investigated using various acyl groups and chemically changing the lysine side chain and the C-terminal of the peptide. A series of N^{ε}-thioacetyllysine compounds were incorporated into a peptide derived from the human p53 protein and α-tubulin. The p53 based peptides showed better inhibition than their α-tubulin counterparts. The most potent SIRT1 inhibitor bears an IC50 value of 180nM, while the most potent SIRT2 inhibitor has an IC50 value of 1.8 uM. Thiomyristoyl peptides, such as BH-TM4 (**66**), were shown to target mainly SIRT2 and SIRT6[170], whereas the thiosuccinyl H3K9TSu (**67**) peptide was synthesised and showed clear selectivity for SIRT5[171]. Several cyclic and linear trifluoroacetyl lysine-containing peptides have also been synthesised and showed selectivity towards SIRT2 and SIRT3 at the low nM affinity level. S2iL5 (**64**) peptide is one of the most potent of these inhibitors[172].



S2iL5 **64** S2iL8 **65** BH-TM4 **66** H3K9TSu **67**

Figure 18 Thioacyllysine-containing compounds.

Other structurally diverse inhibitors have been investigated, in particular suramin (**68**) (Figure 19). Structural studies showed suramin can inhibit SIRT1 by binding in the B and C pockets of the NAD^+ site, and in the substrate active site[173]. Although these findings are encouraging, suramin is known to target many other proteins (e.g. cell surface receptors) and, due to its structure, is likely to have low cell permeability, limiting its utility as a sirtuin inhibitor[174].



Suramin **68**

Figure 19 Structurally diverse inhibitor suramin.

Many other types of compounds have been reported to inhibit sirtuins[175,176] and in the years to come, more and more will be developed. Among the issues with current inhibitors, their potency and selectivity are still very low and there is an urge to improve it and reduce possible off-target effects. Improvement of enzymatic activity assays and *in silico* methods will help understand inhibition potency and sirtuin substrate specificity and design better inhibitors for all sirtuins, instead of mainly class I SIRT1-3.

1.3 Docking and MD in structure-based drug design

The conventional drug development process is incredibly time-consuming and laborious, taking 12-14 years and $1.2-$1.4 billion dollars to get a drug from discovery to market[177]. The introduction of *in silico* methods, during discovery phase, helps in the selection of better and more potent leads, increasing the chances to succeed and save time and resources during the preclinical phase and late clinical trials. Moreover, different computational methods can be used to further study the targets and active sites, analyse their dynamic and kinetic properties, and determine affinity and efficacy of sets of ligands.

1.3.1 *Molecular docking*

One of the approaches to structure-based drug design (SBDD) is the docking of known compounds to a target protein, together with dynamics simulation of molecular complexes such as protein-ligand complexes.

Molecular docking allows the prediction of preferred orientations and main drug interactions (hydrogen bonds, ionic bonds, hydrophobic and van der Waals) between ligand and target, to form a stable complex with less free energy. The predicted binding free energy (ΔG_{bind}) is derived from hydrogen bond (ΔG_{hbond}) and electrostatic contributions (ΔG_{elec}), torsional free energy (ΔG_{tor}), dispersion and repulsion (ΔG_{vdw}), desolvation (ΔG_{desolv}), total internal energy (ΔG_{total}) and unbound system's energy (ΔG_{unb}). A good understanding of the theoretical background can therefore provide important insights into fundamental protein-ligand interactions. There are multiple steps to the molecular docking protocol, from protein and ligand preparation to analysis of the results. In general, each ligand is docked into the receptor, assigned a theoretical binding affinity and ranked accordingly. The best ranked poses are then further evaluated through *in silico* and/or experimental studies. Molecular docking can be achieved through two different strategies, shape complementarity or simulation.

In the shape complementarity approach, both ligand and protein are perceived as surfaces and the complementarity between them helps finding the ligand's binding position on the target. This approach is very quick and can be used for high throughput scanning of ligand libraries.[178,179]

The simulation approach, on the other hand, is a more complex and computationally expensive method. Protein and ligand are separated by a physical distance and the ligand samples the conformational space until finding the position into the protein active site. This involves sampling torsional angle rotations, rotations and translations. After each new conformation, the total energy of the system is calculated. The main advantage of this method is the direct incorporation of ligand flexibility into the calculation. However, this approach is also more time-consuming and expensive. The implementation of a fast optimization method and grid-based tools, made the simulation approach more efficient.[179]

There are several molecular docking tools available (Table 5) that employ different search algorithms such as genetic algorithm, fragment-based algorithms, Monte Carlo and molecular

dynamics algorithms. Within the different types of software, docking can be performed with a rigid target and a flexible ligand (flexible ligand docking), with both target and ligand as rigid molecules (rigid body docking) or with both molecules as flexible (flexible docking).[179,180]

Table 5 List of some successful and widely used docking software, and their algorithms.

Software tool	Search algorithm	Scoring function
Glide (Grid-based Ligand Docking with Energetics)[181]	Monte Carlo	Glide score
AutoDock[182]	Lamarkian genetic algorithm	Empirical free energy function
GOLD (Genetic Optimization for Ligand Docking)[183]	Genetic algorithm	GoldScore, ChemScore
FlexX[184]	Incremental reconstruction	Modified Bohm scoring function
MVD (Molegro Virtual Docking)[185]	Evolutionary algorithm	MolDock score
Fred (Fast Rigid Exhaustive Docking)[186]	Exhaustive search algorithm	Gaussian score function
LigandFit[187]	Monte Carlo	LigScore, Potential Mean Force

Although ligand flexibility is already taken into account in many of the available software, protein flexibility handling is still non-existent or very limited due to the large size and many degrees of freedom. To address this issue, other strategies like soft docking, side-chain flexibility, molecular relaxation, have been implemented[188]. Soft docking is the simplest approach and allows a small degree of overlap between protein and ligand, which is added to the scoring function in the form of interatomic van der Waals energy[189,190]. The side-chain flexibility method uses rotamer libraries to sample the conformational space of the active site side-chains[191]. Finally, molecular relaxation, uses a "dock-and-relax" approach by rigid-body docking the ligand in the active site and then relaxing the protein backbone and side chains around it[192]. This method adds a slight backbone flexibility dimension but also increases significantly the calculation time.

To deal with major structural rearrangements or conformational changes of the protein, more complex methods need to be considered, such as molecular dynamics (MD) simulations, to create an ensemble of protein structures in different conformations.

1.3.2 *Molecular dynamics simulations*

Classical MD simulations allow SBDD protocols that fully account for structural flexibility of the studied complexes, which is essential to predict drug binding and thermodynamic and kinetic properties[193]. MD is an effective tool to predict interaction and motion of proteins by solving Newton's equations of motion for each atom. A force field is used to estimate the forces acting in each atom and, during the simulation, Newton's second law of motion is integrated to generate successive frames of the system. This results in a phase space trajectory in which positions and velocities for each particle in the system are assigned over time.[194] When referring to MD, the chemical behaviour is mainly determined by the nucleus positions and the electron motions are averaged out. In other words, the system can be described using molecular mechanics (MM) energy formulation (Equation 1).

$$U = \sum_{i}^{bonds} \frac{k_{l,i}}{2}(l_i - l_0)^2 + \sum_{i}^{angles} \frac{k_{\theta,i}}{2}(\theta_i - \theta_0)^2 + \sum_{i}^{torsions} \left\{ \sum_{k}^{M} \frac{V_{ik}}{2}[1 - \cos(n_{ik}.\phi_{ik} - \phi_0)] \right\} + \sum_{i}^{pairs} \frac{q_i q_j}{4\pi\varepsilon_i r_{ij}} + \sum_{i,j}^{pairs} \varepsilon_{ij} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right]$$

Equation 1

In this equation, the first three terms refer to the potential energy of intramolecular interactions as a function of bond stretching, bending and torsions. These correspond to the bonded terms of the equation. Bond stretching is described by Hooke's law for a spring while angle bending energies are modelled by polynomial expansions. Both of them are described by harmonic potentials with reference values l_0 and θ_0, and force constants k_l and k_θ.[195] The torsion potential is an expansion of a periodic cosine function that takes into account all 1-4 bond relationships[194]. The fourth and fifth terms represent the non-bonded contributions and correspond to electrostatic and van der Waals interactions, respectively. Electrostatic interactions are accounted by atomic point charges and bond dipole moments and the energy is calculated according to Coulomb's law. The van der Waals interactions are treated with a

12-6 Lennard-Jones potential. The assembly of all these contributions is a force field, the major concern in any MD simulation.

A force field has to be transferable from one molecule to other strictly related molecules, in order to predict the geometry of a new one using derived data to the first one. Thus, a large molecule is no more than the sum of features known for small molecules, combined in different ways. Every force field available is based in equivalent functional equations but each one of them has their own strengths and weaknesses depending on the parameter set.[196]

Due to the high complexity and amount of calculations required to build a space trajectory, MD simulations should be performed on computer clusters or supercomputers, which have hundreds of processors running in parallel. There are multiple molecular dynamics software packages and most of them are compatible with the Message Passing Interface (MPI), a messaging system that allows communication between computers and facilitates the executions of simultaneous tasks.

There are multiple molecular dynamics software packages and each one uses some force field parameters that determine energy functions, bonded and non-bonded interaction specifications, water types and force constraints to be applied. Furthermore, because simulating an infinite number of particles is not feasible, periodic boundary conditions (PBC) are used by the simulation algorithm to mimic a large system. This requires a unit cell with a specific shape that can tile perfectly as a 3D crystal. During the simulation time only the properties of a single cell are recorded and later applied to the images.(74)

Because docking scores suffer from several limitations and, although many advances have been made, sampling during molecular docking still needs to be improved, MD became an important post-processing tool to validate and refine docking structures and predict more accurate binding free energies as well. In this context, MD can provide insights on induced-fit conformational changes, protein-ligand interactions strengthened or weakened from the docking, and structural water molecules in the active site that influence the ligand binding[197]. Many important observables can also be measured from MD trajectories, including internal energy, pressure and heat capacity.[198] One of the most important measurements in drug discovery is the protein-ligand binding free energy, described by the association constant K_a and calculated through Equation 2. In the pharmaceutical research however, the dissociation constant K_d (inversely proportional to K_a) is used more often than K_a. When referring to enzyme inhibitor assays, the inhibitory constant Ki replaces K_d, maintaining the same biochemical interpretation.

$$K_a = \frac{[PL]_{eq}}{[P]_{eq}[L]_{eq}}$$

Equation 2

Where [PL] is the concentration of the protein-ligand complex and [P] and [L] are the concentrations of protein and ligand, respectively.

The connection between experimental equilibrium constants and thermodynamic observables from MD trajectories is provided by the standard free-energy change ΔG_b°, at constant pressure and temperature (Equation 3).

$$\Delta G^\circ = -kT\,ln(K_a C^\circ) = kT\,ln\left(\frac{K_a}{C^\circ}\right)$$

Equation 3

Where C° is a constant defining he standard concentration (typically 1M) and k is the Boltzmann constant.

When the system reaches the equilibrium, the previous equation can be simplified to:

$$\Delta G^\circ = -kTln(K_{eq})$$

Equation 4

Where K_{eq} the equilibrium constant of the reaction.

Different MD-based calculations can be performed to estimate the binding free energy of a protein-ligand complex. Thermodynamic integration (TI) and free energy perturbation (FEP) are two possibilities. However, the exhaustive conformational sampling required makes these approaches highly expensive and limits their application to a few closely related compounds[199]. Two other methods were developed to overcome the previous limitations, including the linear interactions energy (LIE) method[200] and the MM/PBSA method[201].

The LIE method extracts the free binding energy of a compound from simulations of the free and bound state of the ligand[202]. The energy is then calculated based on the electrostatic and van der Waals interactions, according to Equation 5.

$$\Delta G_{bind} = \alpha(V_{bound}^{elec} - V_{(free)}^{elec}) + \beta(V_{bound}^{vdw} - V_{(free)}^{vdw}) + \gamma$$

Equation 5

Where α, β and γ are empirically determined constants and $(V_{bound}^{elec} - V_{free}^{elec})$ and $(V_{bound}^{vdw} - V_{bound}^{vdw})$ represent the average change in electrostatic and van der Waals energy, respectively, of the ligand going from solution to bounding the protein.

This method has been successfully applied to several systems. Osterberg and co-workers[203], studied the binding of different sertindole analogues, hERG K^+ channel blockers. The analogues were docked against a homology model of the open channel and the representative conformations of the most highly populated clusters were submitted to MD simulations. The researchers analysed the protein flexibility and solvent effects, and estimated the binding free energies through LIE, which were in agreement with the experimental values. Similarly, Luzhkov et al.[204] studied the binding of the tetraethylammonium (TEA) ion to the KcsA potassium channel by docking the ligand to the receptor crystal structure. The most stable poses were further submitted to MD and the LIE estimated. Results showed that binding of TEA depends on the number of K^+ ions within the channel, and on the positioning of four tyrosines at the entrance of the pore that stabilize the inhibitor binding.

MM/PBSA combines molecular mechanics and continuum solvent approaches to estimate the binding free energy of a compound.[205,206] After an MD run in explicit solvent, several snapshots are extracted and processed, to remove waters and counter ions, and the total energy of the system is used with:

$$\Delta G_{bind} = \bar{G}_{complex} - [\bar{G}_{protein} - \bar{G}_{ligand}]$$

Equation 6

Where $\bar{G}$ represents the average free energy of the systems and is calculated according to:

$$\bar{G} = \bar{E}_{MM} - \bar{G}_{solvation} - T\bar{S}$$

$$\bar{E}_{MM} = \bar{E}_{int} + \bar{E}_{elec} + \bar{E}_{vdw}$$

$$\bar{G}_{solvation} = \bar{G}_{polar} - \bar{G}_{non-polar}$$

Equation 7

Where E_{MM} is the average MM energy in gas phase, $G_{solvation}$ is the solvation free energy calculated from a Poisson-Boltzmann electrostatic component (G_{polar}) and a solvent-accessible surface area model ($G_{non-polar}$).[207] TS is the entropic contribution and can be omitted if only the relative binding energies of a series of compounds is required. If the compounds are significantly different, absolute binding energies are needed and the entropy term has to be considered.

Multiple studies have used this approach to try and estimate binding energies however, this approach might not be the best since it can produced reasonable energies for some systems but not for others. One example is the work developed by von Langen and colleagues[208] on the study of the human glucocorticoid receptor (hGR) selectivity. After docking of five steroid ligands with FlexX[184], the complexes pose ranking was not in agreement with the experimental affinities. The complexes were submitted to MD and the MM/PBSA binding energy calculated for each complex. The method was able to correctly identify high affinity ligands from low affinity ones however, it could not rank the low affinity ligands correctly.

On the other hand, Kollman and co-workers[209] used a similar approach to predict the binding mode of a HIV-1 reverse transcriptase inhibitor, and successfully estimated both relative and absolute binding energies within a 1.0kcal/mol error. In fact, the most stable binding mode had a -13.2kcal/mol energy, showing good agreement with the -11.6kcal/mol experimental value.

Computational studies have become an essential part of the drug discovery process and many other studies have demonstrated the potential and advantages of using the previously mentioned methods to identify potent drug molecules. Computational approaches add a microscopic perspective to the macroscopic phenomena and provide unique insight into the drug binding mode and affinity, crucial to understand the underlying mechanism of action and therapeutic response. Similar methodology will be used to study the protein-ligand interactions of HDACs along this manuscript.

2. HDAC inhibitors

As already mentioned, several HDAC inhibitors have been approved and are currently used in cancer treatment[69]. Compounds that mimic the substrate binding to HDAC enzymes have been shown to exhibit potent inhibition towards class I HDACs by interacting with both the Zn^{2+} atom in the catalytic site, and the enzyme surface[101].

The results reported in this Chapter were published in a recent paper by our group[210] (Appendix C). In the paper, a series of natural product-derived macrocycles with various zinc-binding groups (ZBGs) were evaluated for HDAC inhibitory potency, binding mode and efficacy on different cell lines. The main findings indicated different potencies and inhibition modes for macrocycles with and without ZBGs.

In order to study the binding affinity and isozyme selectivity of different macrocycles, the ZBGs were excised from a selection of compounds and the resulting norvaline analogues were tested for inhibition against HDAC1-3. Romidepsin, TpxA/B, apicidin A and azumamide C were the most potent for this series. On the other hand, largazole, Cyl-1 and -2, HC-toxin and chlamydocin showed significant losses of potency, indicating low affinity to the enzymes surface and heavy dependency on a strong ZBG to act as inhibitors. Trapoxins A and B, Apicidin A and Cyl-2 scaffolds were then used as scaffolds to reintroduce different ZBGs and consecutive inhibitory activity evaluation. In this case, apicidin A analogues exhibited the highest potencies. Ethyl ketones were more potent than ethyl esters and equipotent to hydroxamic acid-containing analogues. Octyl esters were the least potent analogues.

The binding mode of selected compounds was further studied through competition assays. Carboxylate-containing and ethyl ektone analogues showed competitive inhibition with the substrate while all the norvaline analogues exhibited non-competitive, “mixed” inhibition regarding the inhibitor concentration and competitive inhibition towards the substrate. To further investigate the reason for the “mixed” inhibition mode, a trapoxin B ethyl ketone and a trapoxin B norvaline analogues were used to perform an exclusivity study and measure the steady-state kinetics of HDAC3 inhibition. The results from this experiment showed a mutually exclusive binding for the two inhibitors, which means norvaline analogs possibly bind in the same pocket as the substrate and other inhibitors with a ZBG.

However, the specifics of how the binding occurs and how the analogues are positioned in the pocket cannot be uncovered with the experiments reported. For this reason, I used computational approaches to examine a selection of compounds that includes hydroxamic acid-containing and norvaline inhibitors. In this chapter I report the molecular modelling

study of a selected set of HDAC inhibitors (Figure 20). The goal is to rationalize compound potencies by understanding how amino acid composition and stereochemistry can influence their affinity to HDAC2 and HDAC3. This chapter is organized in sections as follows: the first part reports the conformational studies performed on the macrocycle compound selection. The second part includes the study of the binding mode of each compound to HDAC2 and HDAC3, as well as their affinities. Finally, the third section reports and discusses the structural and dynamic results from the simulations with the inhibitors bound to either HDAC2 or HDAC3.

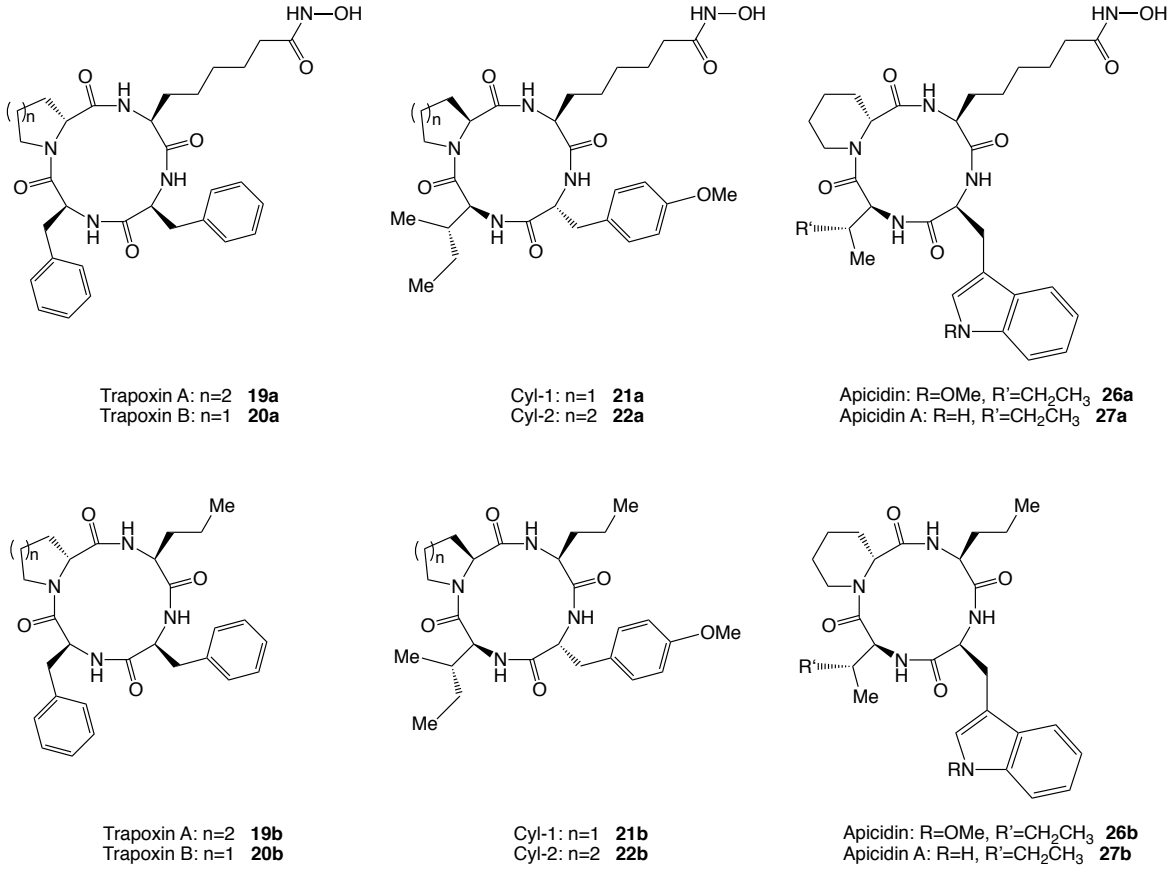


Figure 20 Macrocyclic ligand library. The 12 inhibitors were used to study HDAC2 and 3.

2.1 Macrocycle conformational analysis

Conformational preferences of TpxA (**19a**), TpxB (**20a**), Cyl-1 (**21a**), Cyl-2 (**22a**), Api (**26a**) and ApiA (**27a**) analogues were investigated using two different approaches. First, using the macrocycle conformational sampling algorithm[32] within MacroModel[33], and second using a "sawtooth" simulated annealing (SA) NVT protocol in water and in DMSO (for details on the methods see Appendix A). The selected compounds were built in Maestro[211] through

mutation of experimentally solved structures. Trapoxins A/B, apicidin and apicidin A analogues were built using trapoxin A[212] and apicidin[213] X-ray structures as templates, respectively, while Cyl-1 and Cyl-2 analogues were built using a Cyl-1 NMR solved structure[214] as reference. The stereochemistry of each compound was maintained as shown in Figure 20.

2.1.1 MacroModel Conformational Search

This conformational search method was parameterized using published macrocycle structures from Protein Data Bank (PDB) and Cambridge Structural Database (CSD), making it more accurate for this type of compound. The search was conducted with implicit water solvent. Table 6 shows all possible combinations of conformations, and the number of conformers in each one.

Table 6 Statistics from Macrocycle Conformational Sampling for hydroxamic acid containing analogues. Torsion angles between -90° and 90° were considered *cis*, torsion angles outside this interval were considered *trans*.

Conformation	Number of occurrences					
	TpxA	**TpxB**	**Cyl-1**	**Cyl-2**	**Api**	**ApiA**
TTTT	1937	1849	560	437	1581	1361
TTTC	390	220	33	210	60	191
TTCT	148	293	25	36	90	87
TTCC	38	38	6	9	1	10
TCTT	707	282	1579	2099	1191	1606
TCTC	150	42	528	648	64	145
TCCT	88	46	170	166	18	29
TCCC	9	25	48	34	3	7
CTTT	111	120	23	57	93	81
CTTC	95	11	1	16	6	9
CTCT	27	29	17	4	4	14
CTCC	9	11	2	1	0	0
CCTT	65	29	51	43	58	79
CCTC	17	16	32	28	10	18
CCCT	11	10	25	17	4	6
CCCC	4	7	13	4	0	3
Total	3806	3028	3113	3809	3183	3646
Lowest energy	TTTT	TTTT	TCTT	TCTT	TTTT	TCTT

In the case of ligands with Zn^{2+} binding groups, the most frequent conformation is all-*trans* for TpxA, TpxB and Api whereas Cyl-1, Cyl-2 and ApiA show preference for conformations with a *cis* tertiary amide. The lowest energy structure is included in the most frequent conformation category for all ligands except for apicidin, where the lowest energy conformer is *TCTT.*

The lowest energy conformers for trapoxin A and B exhibit all-*trans* conformation, aligning all three NH groups on the same side of the ring (Figure 21). The only difference between these conformers is a slightly distinct orientation of their side chains. In the case of Cyl-1/2 and apicidin and apicidin A, the lowest energy conformers contain a tertiary amide in the *cisoid* conformation. In Cyl-1 and -2 this results in the alignment of only two NHs in one side of the ring and three carbonyl groups on the other side, promoting hydrogen bonding between the carbonyl of Pro and the NH of Ile. In apicidin and apicidin A, all four carbonyls align on the same side of the backbone plane. This allows for hydrogen bonding between the H protons and the hydroxamic acid. The next conformers however, exhibit four *trans* amide bonds, with all three NHs aligned on the same plane, allowing the carbonyl of Ile to form an hydrogen bond with the NH of the TrpOMe or Asuha (in apicidin or apicidin A, respectively). Similarly to trapoxin A/B, the macrocycle backbone conformation is identical between Cyl-1 and Cyl-2, and Api and ApiA, with only slight differences in the side chains.

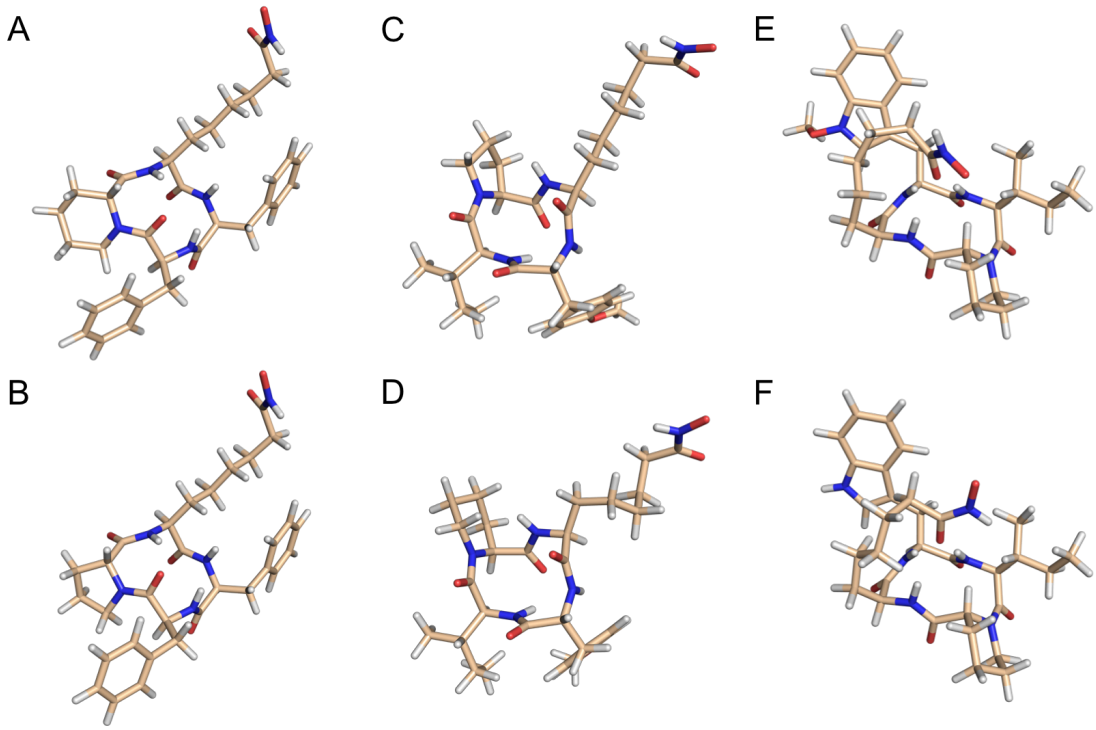


Figure 21 Lowest energy conformers of TpxA (A), TpxB (B), Cyl-1 (C), Cyl-2 (D), Api (E) and ApiA (F) hydroxamic acid analogues. Carbon atoms are shown in wheat, oxygens in red, nitrogens in blue and hydrogens in white.

Similar findings are reported for the norvaline analogues, with the exception of apicidin and apicidin A. In the case of these analogues, the lowest energy structures of TpxA/B and Api/ApiA contain all-*trans* amide bonds, whereas Cyl-1/2 contain a *cis*-arrangement at the tertiary amide (Figure 22). All the lowest energy structures are members of the most frequent conformational cluster (Table 7).

Table 7 Statistics from Macrocycle Conformational Sampling for norvaline analogues. Details see **Table 6** caption.

Conformation	Number of occurrences					
	TpxA	**TpxB**	**Cyl-1**	**Cyl-2**	**Api**	**ApiA**
TTTT	427	223	171	234	655	539
TTTC	150	129	101	103	143	170
TTCT	146	117	120	75	123	140
TTCC	67	65	64	33	9	44
TCTT	368	149	286	541	529	424
TCTC	146	114	192	231	128	160
TCCT	125	75	103	151	92	81
TCCC	44	56	77	51	24	18
CTTT	130	82	84	99	116	107
CTTC	33	58	43	46	29	28
CTCT	72	78	57	39	20	57
CTCC	21	37	18	20	4	32
CCTT	126	133	81	84	89	148
CCTC	66	94	67	55	33	61
CCCT	52	65	54	65	24	27
CCCC	21	27	25	22	14	15
Total	1994	1502	1543	1849	2032	2051
Lowest energy	TTTT	TTTT	TCTT	TCTT	TTTT	TTTT

In the lowest energy conformation, both Tpx A/B and Api/ApiA exhibit all three NH protons aligned on the same side of the ring. The orientation of the carbonyl of the cyclic residue allows hydrogen bonding with the NH of the norvaline residue. On the other hand, Cyl-1/2 contain a *cis* tertiary amide and the hydrogen bond is established between the cyclic residue carbonyl and the Ile NH.

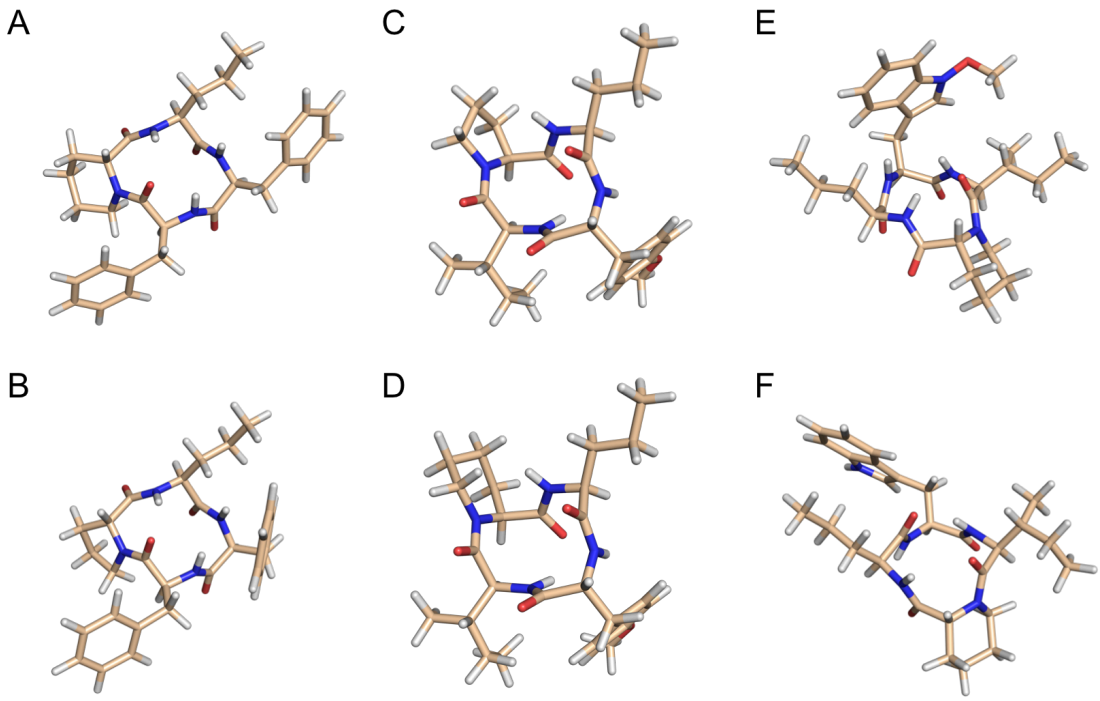


Figure 22 Lowest energy conformers of TpxA (A), TpxB (B), Cyl-1 (C), Cyl-2 (D), Api (E) and ApiA (F) norvaline analogues. Details as in **Figure 21**.

2.1.2 Simulated Annealing

In order to study the behaviour of the ligands in water and in DMSO, each compound was solvated in a 10Å box of water and in a 10Å box of DMSO, totalling 12 systems. Each system was gradually heated to 565K and 675K in water and in DMSO, respectively. The temperature was then slowly decreased over 100ns so a better sampling of the conformational space could be achieved. The last frame of each Simulated Annealing simulation was further subjected to MD at 300K.

For each SA simulation, the torsion angles of the amide bonds in the macrocycle were analysed for the most frequent populations (Figure 23). As expected, in both solvents, the most flexible amide bond is the tertiary amide. Trapoxin A can assume both *cis* and *trans* conformations at the tertiary amide, in water and in DMSO, while TpxB most frequently adopts an all-*trans* conformation in both. In the case of Cyl-1 and Cyl-2 the tertiary amide can assume a *trans*-arrangement but is much more frequent in the *cisoid* conformation in both solvents. On the other hand, the conformational preference of apicidin and apicidin A varies depending on the solvent. A *cis* conformation at the tertiary amide is more frequent for

apicidin in water and apicidin A in DMSO. The opposite is also true for the *trans* tertiary amide arrangement.

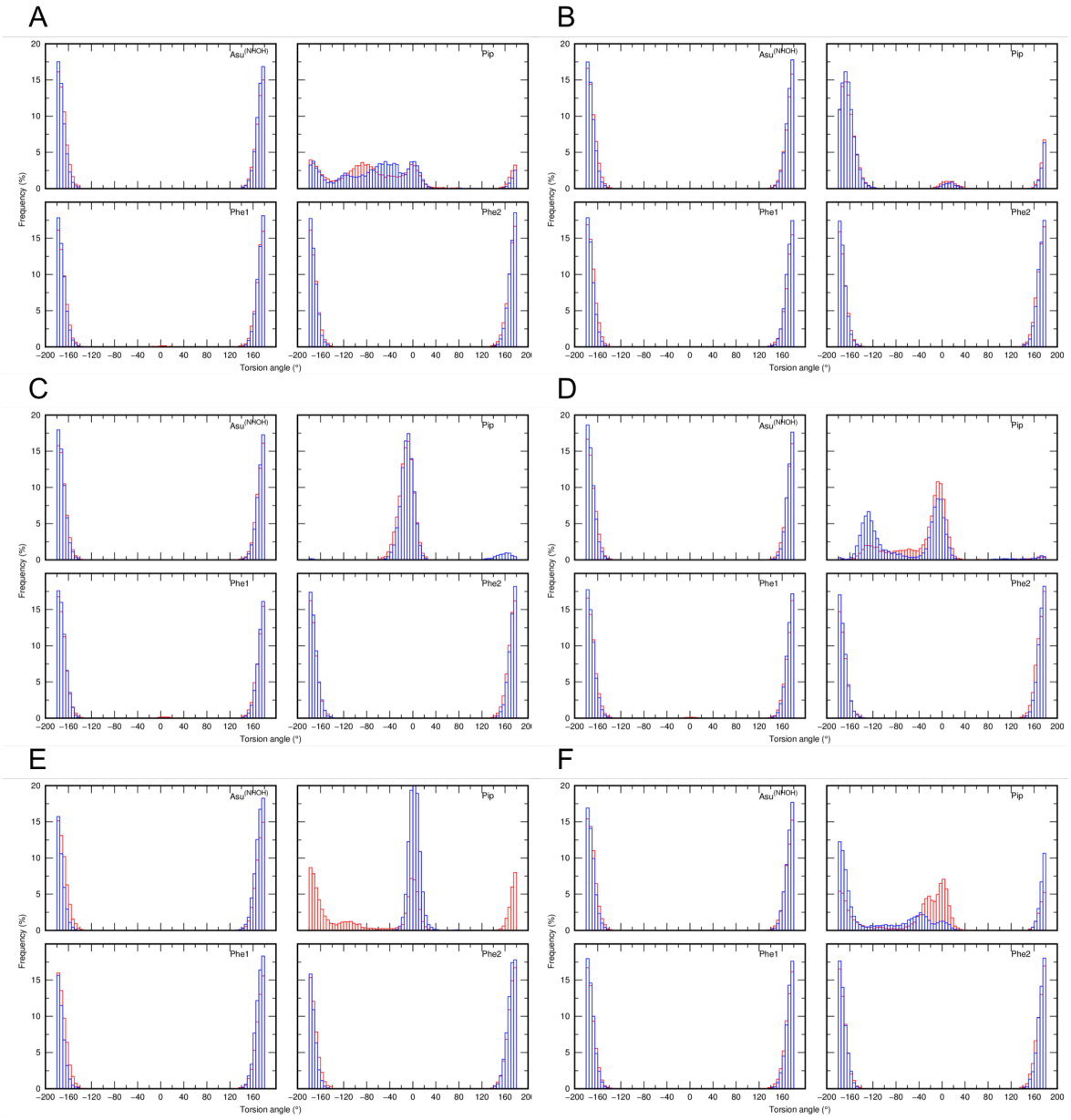


Figure 23 Histograms of the torsion angles of the amide bonds in the macrocycle backbone. Measurements for hydroxamic acid TpxA (A), TpxB (B), Cyl-1 (C), Cyl-2 (D), Api (E) and ApiA (F) in water systems are represented in blue and DMSO systems in red.

2.1.3 DFT optimization

Each ligand, in an all-*trans* conformation and in a *TCTT* conformation, was fully optimized with DFT/6-31G**. Cyl-2 was also modified, so all *trans* amides and all NH/CH-*alpha* proton pairs would be *anti* to each other, and further optimized. The table below shows the energies for each ligand with the two different conformations.

Table 8 DFT energies. SA initial structures were used as input. Calculations performed in water implicit solvent.

Conformation	Compound	Energy (hartree)	Energy (kcal/mol)
TTTT	TpxA_NHOH	-1968.848528	-1235470.17
	TpxA_Nva	-1646.838754	-1033406.14
	TpxB_NHOH	-1929.544223	-1210806.37
	TpxB_Nva	-1607.525824	-1008736.92
	Cyl-1_NHOH	-1930.941351	-1211683.08
	Cyl-1_Nva	-1608.929389	-1009617.67
	Cyl-2_NHOH	-1970.252707	-1236351.31
	Cyl-2_NHOH_NH_CH_pairs	-1895.600699	-1189506.50
	Cyl-2_Nva	-1648.241603	-1034286.44
	Cyl-2_Nva_NH_CH_pairs	-1648.247891	-1034290.39
	Api_NHOH	-2101.776542	-1318883.70
	Api_Nva	-1779.769138	-1116821.15
	ApiA_NHOH	-2062.449275	-1294205.48
	ApiA_Nva	-1740.456759	-1092152.28
TCTT	TpxA_NHOH	-1968.839538	-1235464.53
	TpxA_Nva	-1646.827506	-1033399.08
	TpxB_NHOH	-1929.532405	-1210798.95
	TpxB_Nva	-1607.509819	-1008726.88
	Cyl-1_NHOH	-1930.938698	-1211681.41
	Cyl-1_Nva	-1608.93624	-1009621.97
	Cyl-2_NHOH	-1970.25757	-1236354.36
	Cyl-2_Nva	-1648.251222	-1034292.48
	Api_NHOH	-2101.767179	-1318877.82
	Api_Nva	-1779.751527	-1116810.10
	ApiA_NHOH	-2062.458571	-1294211.32
	ApiA_Nva	-1740.428896	-1092134.80

In both conformations, ligands with the ZBG are lower in energy than their analogues without ZBG. Both analogues of TpxA and TpxB seem to be more favourable with all-*trans* amides, the same way as Cyl-1 Nva and Cyl-2 seem more favourable with a *cis* tertiary amide. Cyl-2 Nva

with *anti* NH/CH-*alpha* protons is slightly more favourable in energy than its analogue with all NH protons on the same side of the macrocycle.

The torsion angles of the macrocycle amide bonds measured in both approaches show a good agreement between them and with the experimental data (see Table 9). In general, Cyl-1/2 assume a TCTT conformation whereas trapoxins and apicidins preferentially exhibit TTTT conformations. However, the tertiary amide proved to be highly flexible, visiting both *trans* and *cis* conformations in all of the theoretical experiments. The preferred conformations proved to be dependent not only on the solvent but also on the presence or lack of a ZBG and the physical state (solid or solution) of the sample. One example of such is the conformation of TpxA, which, on the solid state, contains a *cis* tertiary amide necessary for hydrogen binding to other monomers. When the exact same monomer is isolated and energy minimized, the tertiary amide assumes a lower energy *trans* arrangement.

Table 9 Summary of compound conformational preferences.

		Experimental	Conf search		SA	
			Lowest energy	Most frequent	Water	DMSO
TpxA	Asuha	TCTT	TTTT	TTTT	TTTT/TCTT	TTTT/TCTT
	Nva	-	TTTT	TTTT		
TpxB	Asuha	-	TTTT	TTTT	TTTT	TTTT
	Nva	-	TTTT	TTTT		
Cyl-1	Asuha	-	TCTT	TCTT	TCTT	TCTT
	Nva	-	TCTT	TCTT		
Cyl-2	Asuha	TCTT	TCTT	TCTT	TCTT	TCTT
	Nva	-	TCTT	TCTT		
Api	Asuha	TTTT	TCTT	TTTT	TCTT	TTTT
	Nva	-	TTTT	TTTT		
ApiA	Asuha	-	TCTT	TCTT	TTTT	TCTT
	Nva	-	TTTT	TTTT		

2.2 Docking

Molecular docking has been used as one of the most common methods for structure-based drug design since the early 1980s[215]. This approach allows the thorough study and characterization of protein-ligand interactions at the atomic level, providing valuable insight into their behaviour in the binding site. In a first instance, the ligand conformation and its position and orientation in the binding site is predicted using a sampling algorithm; the resulting poses are then assessed for binding affinity and ranked *via* a scoring function.

2.2.1 Protocol validation

The docking protocol was validated by comparing protein interactions and RMSD values of docked ligands with the corresponding co-crystalized structures. Crystal structures of HDAC1, -2 and -3 were downloaded from the Protein Data Bank and each ligand was redocked and the docking score calculated (Table 10).

Table 10 Protocol validation results with Glide.

	PDB ID	XP PoseRank	XP GScore	Glide emodel	RMSD to input
HDAC1	4BKX	1	-1.868	-19.239	0.538
	5ICN	1	-12.02	-109.92	8.398
		2	-11.831	-114.921	8.418
		3	-11.796	-114.578	8.719
		4	-11.145	-113.882	9.794
		5	-10.691	-108.691	9.655
		6	-10.174	-110.234	9.691
		7	-10.069	-108.612	3.735
		8	-9.636	-111.913	6.778
HDAC2	3MAX_A	1	-14.414	-105.097	0.227
	3MAX_B	1	-13.138	-103.342	0.438
	3MAX_C	1	-14.931	-108.03	0.277
	4LXZ_A	1	-11.015	-85.547	0.676
		2	-10.68	-88.526	0.725
		3	-10.628	-84.433	0.861
		4	-10.582	-84.294	0.683
		5	-10.45	-85.604	0.7
	4LXZ_B	1	-11.278	-81.627	0.457
		2	-11.121	-84.005	0.556
		3	-11.089	-80.615	0.592
		4	-10.887	-82.988	0.303
		5	-10.811	-82.679	0.685
		6	-9.889	-80.777	0.776
	4LXZ_C	1	-11.134	-85.339	0.588
		2	-10.735	-83.597	0.157
		3	-10.659	-85.91	0.6
		4	-9.806	-82.178	0.839
	4LY1_A	1	-14.673	-127.075	0.402
		2	-13.728	-126.055	0.902
	4LY1_B	1	-13.794	-132.118	0.432
	4LY1_C	1	-15.599	-147.358	0.281
HDAC3	4A69_A	1	-3.45	-29.933	0.468

	PDB ID	XP PoseRank	XP GScore	Glide emodel	RMSD to input
HDAC1	4BKX	1	-1.868	-19.239	0.538
	5ICN	1	-12.02	-109.92	8.398
		2	-11.831	-114.921	8.418
		3	-11.796	-114.578	8.719
		4	-11.145	-113.882	9.794
		5	-10.691	-108.691	9.655
		6	-10.174	-110.234	9.691
		7	-10.069	-108.612	3.735
		8	-9.636	-111.913	6.778
HDAC2	3MAX_A	1	-14.414	-105.097	0.227
	3MAX_B	1	-13.138	-103.342	0.438
	3MAX_C	1	-14.931	-108.03	0.277
	4LXZ_A	1	-11.015	-85.547	0.676
		2	-10.68	-88.526	0.725
		3	-10.628	-84.433	0.861
		4	-10.582	-84.294	0.683
		5	-10.45	-85.604	0.7
	4LXZ_B	1	-11.278	-81.627	0.457
		2	-11.121	-84.005	0.556
		3	-11.089	-80.615	0.592
		4	-10.887	-82.988	0.303
		5	-10.811	-82.679	0.685
		6	-9.889	-80.777	0.776
	4LXZ_C	1	-11.134	-85.339	0.588
		2	-10.735	-83.597	0.157
		3	-10.659	-85.91	0.6
		4	-9.806	-82.178	0.839
	4LY1_A	1	-14.673	-127.075	0.402
		2	-13.728	-126.055	0.902
	4LY1_B	1	-13.794	-132.118	0.432
	4LY1_C	1	-15.599	-147.358	0.281
	4A69_B	1	-3.227	-26.374	0.254

All ligands were positioned very similarly to the crystal structures, mostly interacting with the same residues in the pocket and on the enzymes surfaces. Among these residues, Tyr303/308/298, His140/134/145, His141/135/146 and Phe205/155/144 seem to favour the binding of ligands inside the pocket. On the other hand, interactions with His28/33/22 and Asp99/104/93 seem to stabilize the cap groups on the enzyme surface. Interestingly,

ligands in entries 5ICN and 4LXZ closely relate to the HDAC inhibitor pharmacophore in Chapter 1 and are shown to interact with these residues.

The RMSD values calculated with respect to the co-crystallized ligands were all below 1Å, proving this protocol to be accurate and reliable. The only exception was ligand 5ICN. In this case, the RMSD values were substantially increased by the distinct orientation of the cap group on the enzyme surface. However, the linker and the hydroxamate are correctly positioned in the pocket showing that the predicted binding mode was still correct.

Although the docking scores calculated for the ligands cannot be used as absolute measurements of binding affinity, they successfully identified stronger and weaker binders as confirmed by the scores below -10/-12 (Schrodinger support Article ID: 639 - Last Modified: December 4, 2010 - 12:00am).

2.2.2 Macrocycle docking

Once the docking protocol was properly validated, we proceeded to the docking of the trapoxins A/B, Cyl-1/2, apicidin and apicidin A analogues. For this matter, two high-resolution X-ray crystal structures were chosen for the docking studies, HDAC2 (PDBID: 4LY1) and HDAC3 (PDBID: 4A69). Docking was performed using Glide[181]. Grids suitable for peptide docking were generated for each receptor using the co-crystallized acetate as the centre of the grid. Docking was performed in extra precision (XP) mode, with flexible ligand and post-docking minimization.

The docked poses of the hydroxamic acid analogues (Figure 24) showed a similar binding mode in both HDAC2 and 3. The zinc binding group is positioned next to the zinc ion with which they are expected to interact, and the cap group (macrocycle) sits on the enzyme surface.

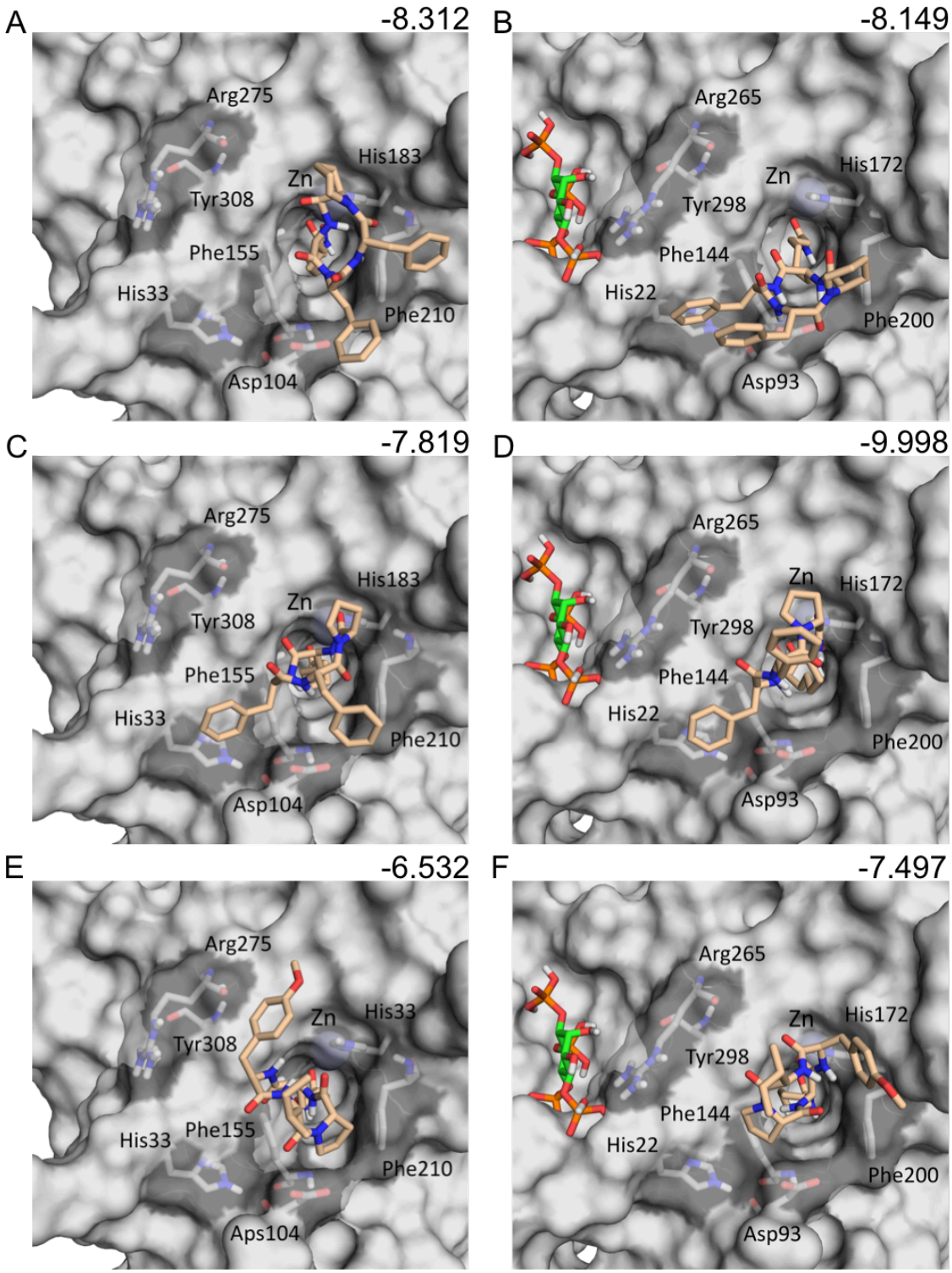


Figure 24 Docking of hydroxamic acid compounds TpxA (A, B), TpxB (C, D) and Cyl-1 (E, F) to HDAC2 (left column) and HDAC3-SMRT (right column). Enzyme is represented as the grey surface, residues are shown as sticks with grey carbons, tested compounds are shown as sticks with wheat carbons and insositol phosphate is shown as sticks with green carbons. Values no the right upper corner correspond to the docking scores.

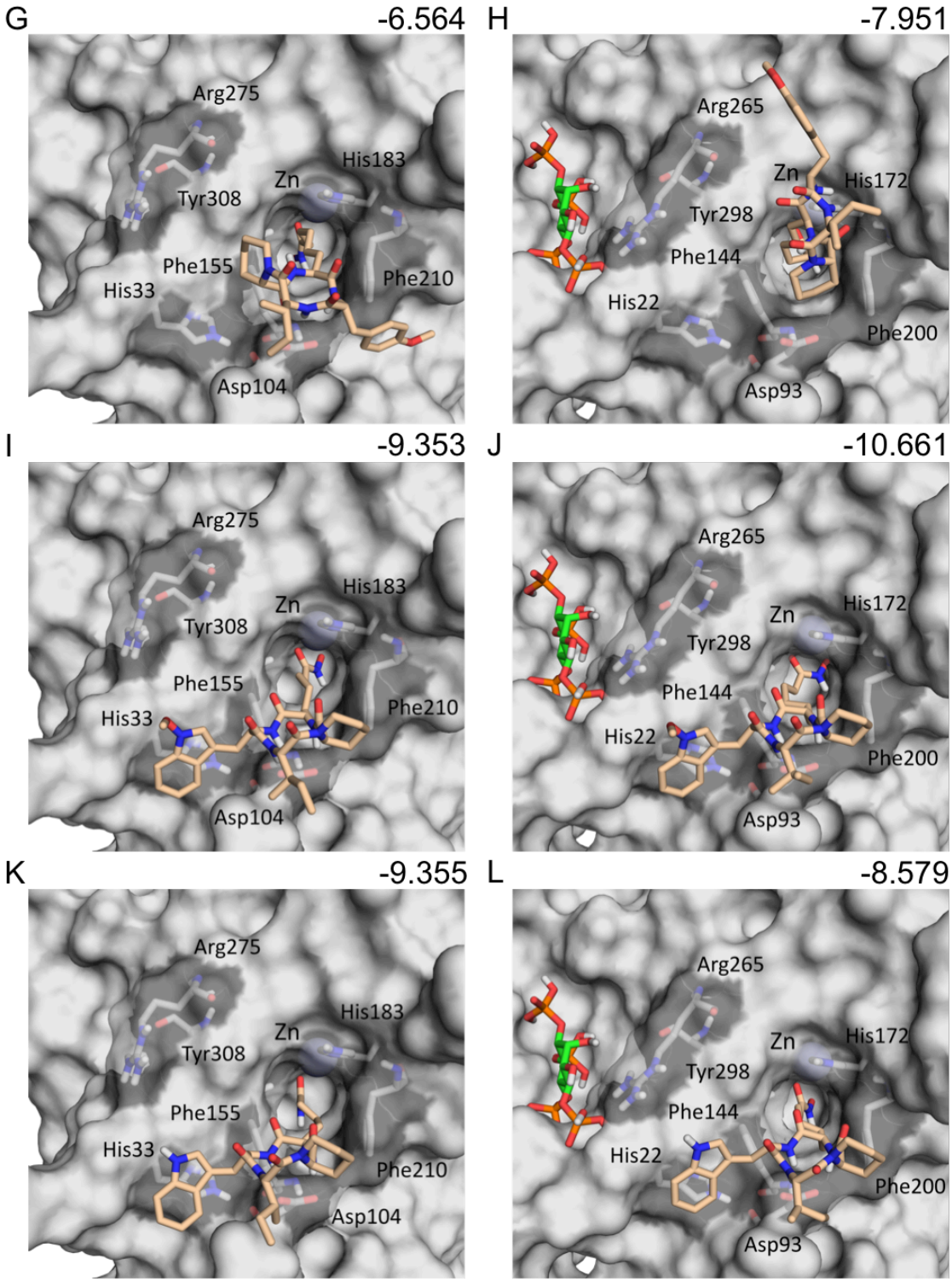


Figure 24 (continued) Docking of hydroxamic acid compounds Cyl-2 (G, H), Api (I, J) and ApiA (K, L) to HDAC2 (left column) and HDAC3-SMRT (right column).

Apicidin and apicidin A analogues interact with His33/22 by establishing π-π stacking interactions through the aromatic side chain of Trp(OMe/Me), while also interacting with the carboxylate side chain of Asp104/93 through hydrogen bonding (backbone NH amide groups), which is in agreement with previous studies on azumamides[216]. Trapoxin A and B can also establish hydrogen bonds with Asp104/93 and hydrophobic interactions with His33/22 but not both in one pose, which might decrease their binding affinity. However, this hindrance seems to be compensated by an increased number of molecular interactions with residues inside the binding pocket such as, Tyr308/298, His183/172, His145/134 and His146/135. On the other hand, Cyl-1 and Cyl-2 are positioned in a different manner at the enzyme surface such that interactions with both Asp104/93 and His33/22 residues cannot be established. Moreover, few interactions are established with residues inside the pocket. These findings, together with the docking scores, indicate a weaker binding affinity for Cyl-1 and Cyl-2 and, a higher binding affinity for apicidin and apicidin A.

On the other hand, the norvaline analogues showed a completely different binding mode, projecting an aromatic side chain towards the binding pocket (Figure 25). In this case, the docking scores were much lower than for the equivalent hydroxamic acid analogues, except for Cyl-1 and Cyl-2. Surprisingly, these analogues showed the higher docking scores among the norvaline analogues, which we believe be due to the unique interaction between the methoxy group of Phe and the zinc ion.

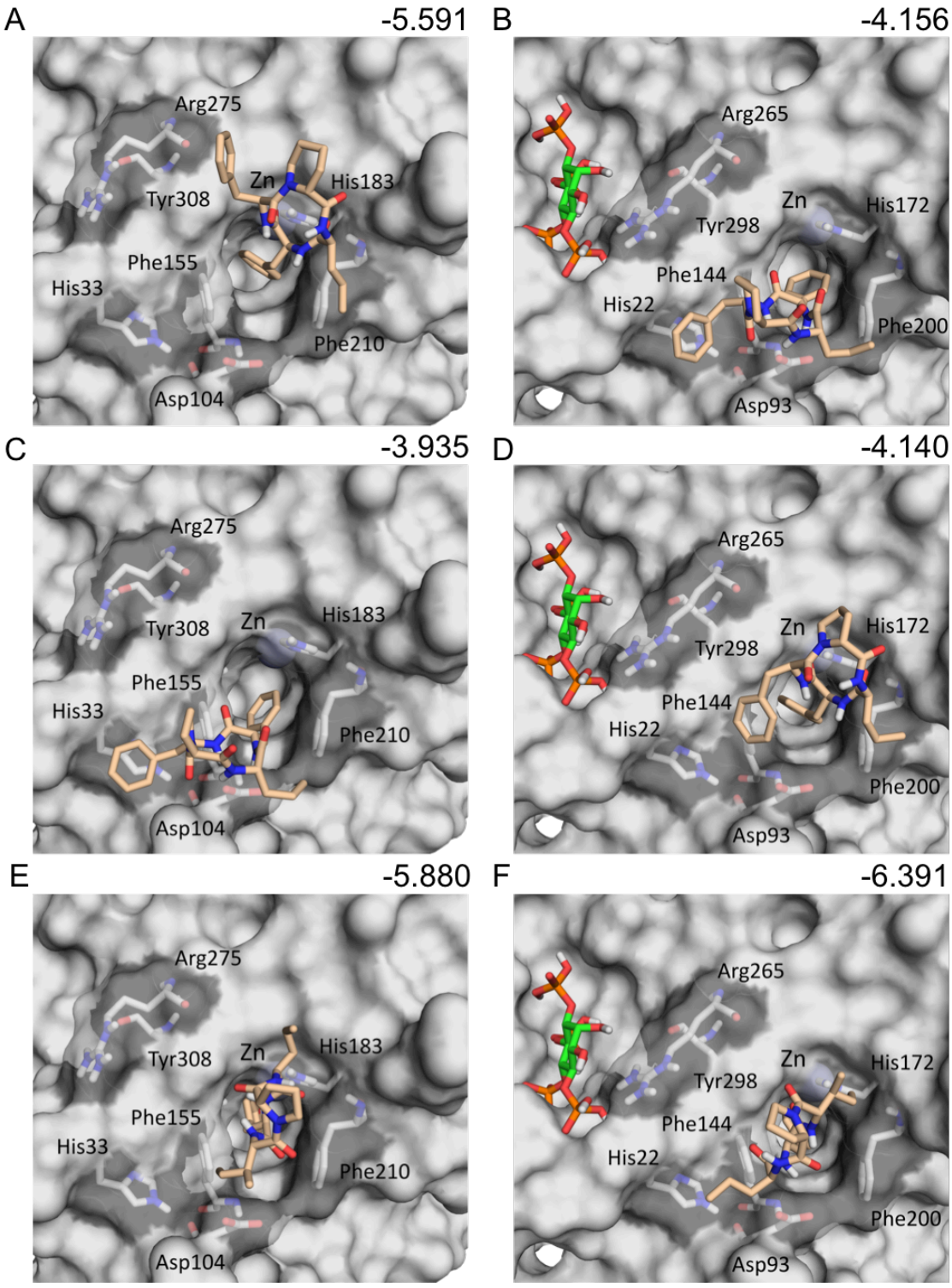


Figure 25 Docking of norvaline analogues: TpxA (A, B), TpxB (C, D) and Cyl-1 (E, F) to HDAC2 (left column) and HDAC3-SMRT (right column). Details as in **Figure 24**.

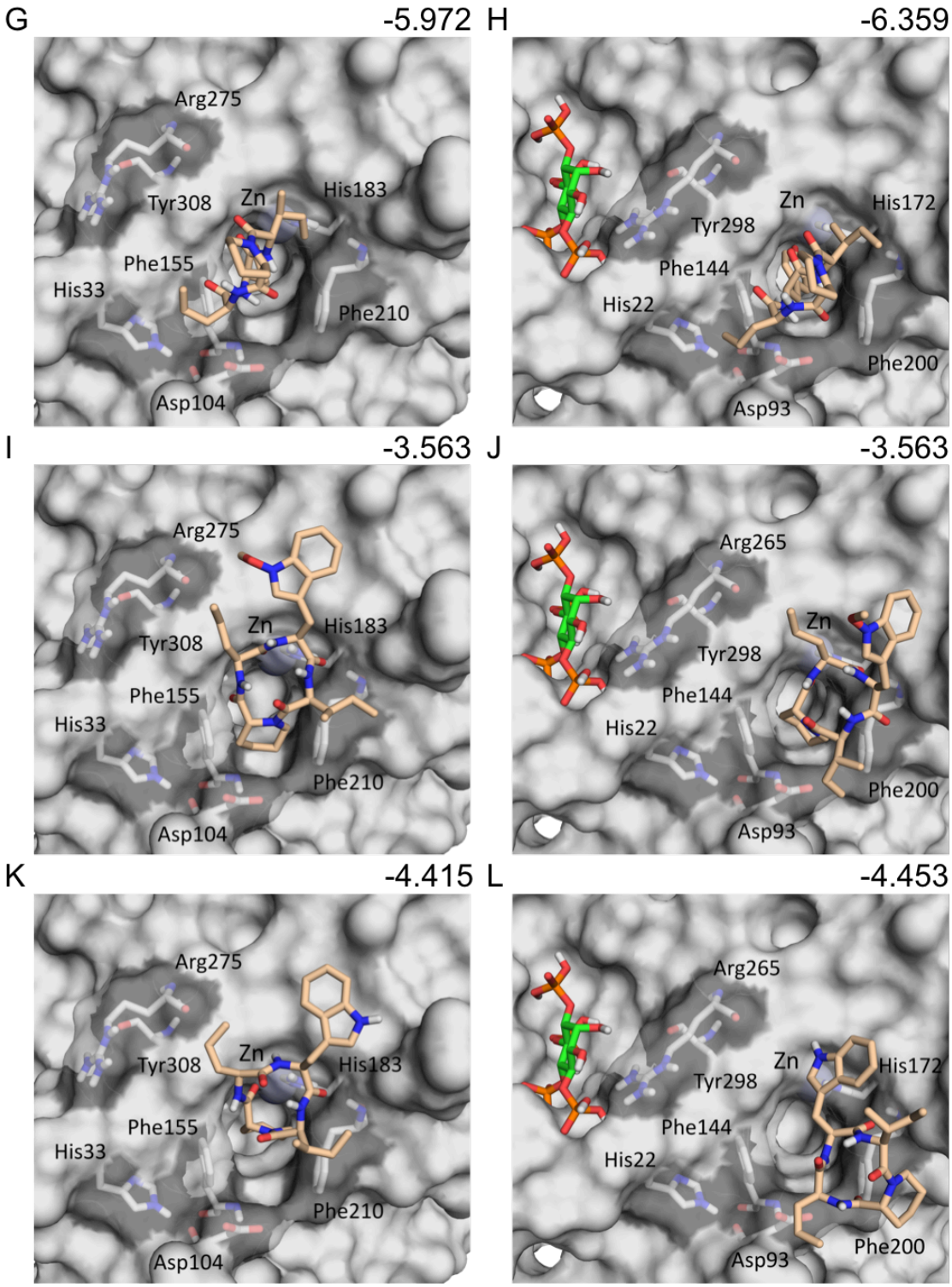
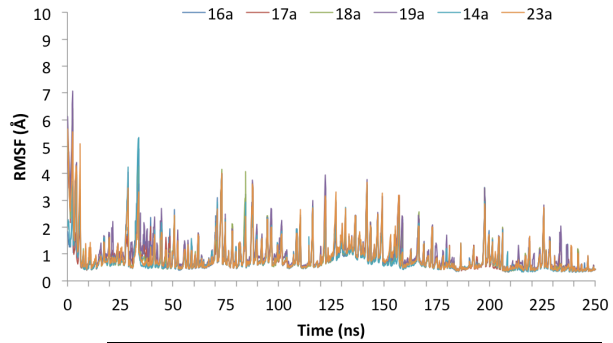


Figure 25 (continued) Docking of norvaline analogues Cyl-2 (G, H), Api (I, J) and ApiA (K, L) to HDAC2 (left column) and HDAC3-SMRT (right column). Details as in **Figure 24**.

2.3 MD simulations

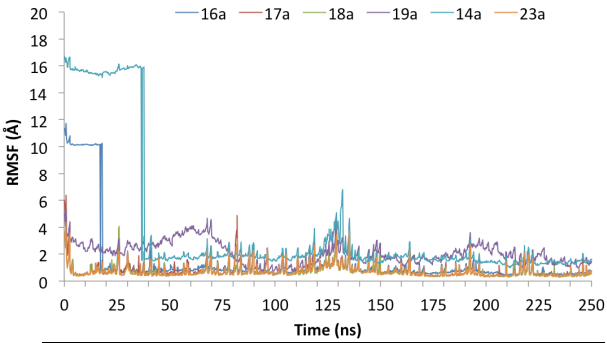
In order to further investigate interactions between analogues and enzymes, the best pose from each docking study was used as input for a 250ns MD simulation run (see details in Appendix A). Analysis of protein heavy atoms RMSF during 250ns (Figure 26) showed higher values in the beginning of the simulation. These tend to decrease and stabilise below ≈3Å, which indicates the protein tends to stabilise and the system equilibrates.

A



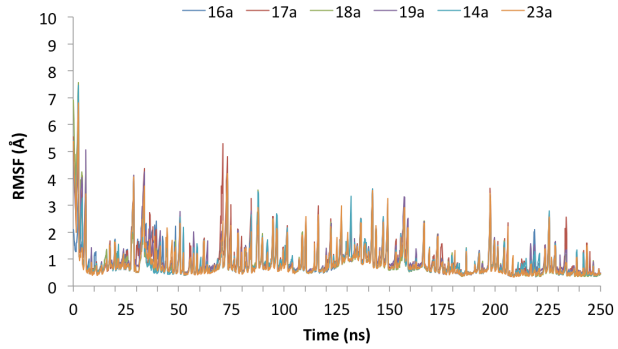
	16b	17b	18b	19b	14b	23b
Average	0.63	0.70	0.57	0.67	0.63	0.58
Std. Dev.	0.35	0.38	0.32	0.31	0.32	0.31
Min	0.35	0.38	0.34	0.35	0.34	0.34
Max	2.72	2.56	2.68	2.53	2.80	2.55

B



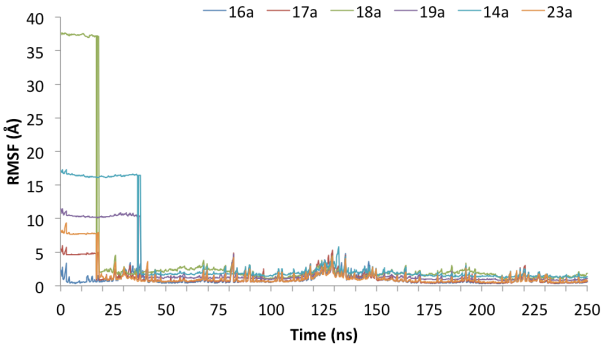
16b	17b	18b	19b	14b	23b
0.65	0.72	1.35	0.99	1.44	0.75
0.39	0.48	0.29	0.23	0.23	0.38
0.32	0.36	0.86	0.76	1.17	0.43
2.13	3.05	2.32	2.20	2.56	2.24

C



	16a	17a	18a	19a	14a	23a
Average	0.61	0.57	0.58	0.69	0.54	0.59
Std. Dev.	0.34	0.31	0.31	0.36	0.30	0.32
Min	0.36	0.33	0.34	0.37	0.32	0.35
Max	2.68	2.69	2.60	2.82	2.66	2.74

D



16a	17a	18a	19a	14a	23a
0.61	0.57	0.58	0.69	0.54	0.59
0.34	0.31	0.31	0.36	0.30	0.32
0.36	0.33	0.34	0.37	0.32	0.35
2.68	2.69	2.60	2.82	2.66	2.74

Figure 26 Protein RMSF calculated for heavy atoms over 250ns. Left column shows the results for HDAC2 while the right column represents HDAC3. **A** and **B** correspond to hydroxamic acid analogues and **C** and **D** norvaline analogues. Statistic values on tables are given in Å and were calculated for the last 50ns of simulation time.

During the simulations, the stability of each inhibitor in complex with HDAC2 and HDAC3 was investigated and their binding affinities evaluated through analysis of the MD trajectories and protein-ligand interactions established. During the first 100ns of simulation in HDAC2 (Figure 27a) TpxA^{Asuha} cap group is positioned slightly upwards the pocket entrance due to hydrogen bonding with Phe210. This interaction is then interrupted and the downward movement of the macrocycle allows hydrogen bonding with Asp104, and π-π stacking between Phe aromatic side chain and Tyr308 aromatic ring (Figure 27a 250ns). In HDAC3, TpxA^{Asuha} was even more stable showing minor rotations at the macrocycle level (Figure 28). Hydrogen bonds between the macrocycle amide NH groups and the carboxylate side chain of Asp104, as well as π-π stacking interactions between the macrocycle Phe aromatic side chain and His22 side chain, were established throughout the entirety of the simulation.

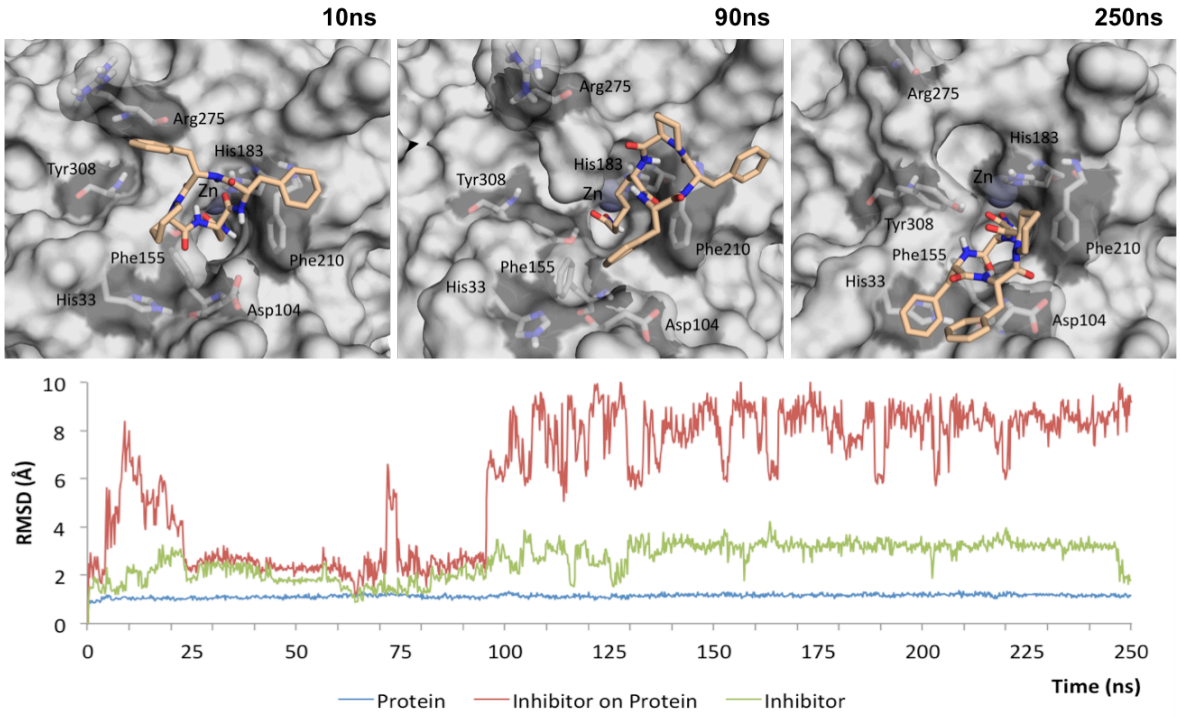


Figure 27 MD simulation snapshots of TpxA^{Asuha} in HDAC2 and RMSD variation throughout 250ns of simulation. Enzyme in snapshots is represented as the grey surface, protein residues are shown as sticks with grey carbons and studied compounds are shown as sticks with wheat carbons. RMSD for the protein backbone and the studied compound are represented by the blue and green lines, respetively, and red line corresponds to the compounf RMSD in relation to its position in the enzyme.

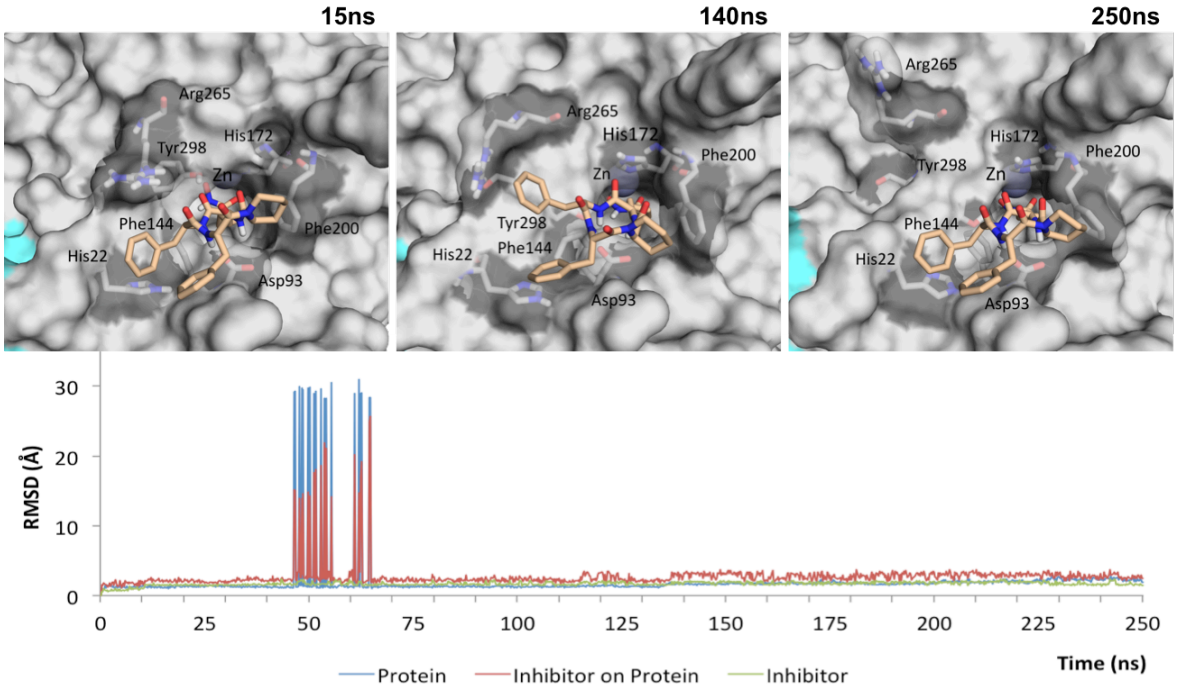


Figure 28 MD simulation snapshots of TpxA^{Asuha} in HDAC3 and RMSD variation throughout 250ns of simulation. Enzyme in snapshots is represented as the grey surface, protein residues are shown as sticks with grey carbons and studied compounds are shown as sticks with wheat carbons. Co-repressor is shown as blue surface. RMSD for the protein backbone and the studied compound are represented by the blue and green lines, respetively, and red line corresponds to the compounf RMSD in relation to its position in the enzyme.

The analysis of the RMSD during TpxB^{Asuha} simulation suggests this inhibitor produces more stable complexes than the previous TpxA^{Asuha} system (Figure 29, Figure 30). However, interactions between the macrocycle and the enzyme surface seem to be slightly more limited. In HDAC2 (Figure 29), these interactions include occasional water bridges with Asp104, His183 and Phe210, and hydrophobic interactions with Tyr308 (π-π stacking) and Arg275 (π-cation). On the other hand, in HDAC3, TpxB^{Asuha} established hydrogen bonds with His172 and Phe200, which position the macrocycle away from Asp93 (Figure 30).

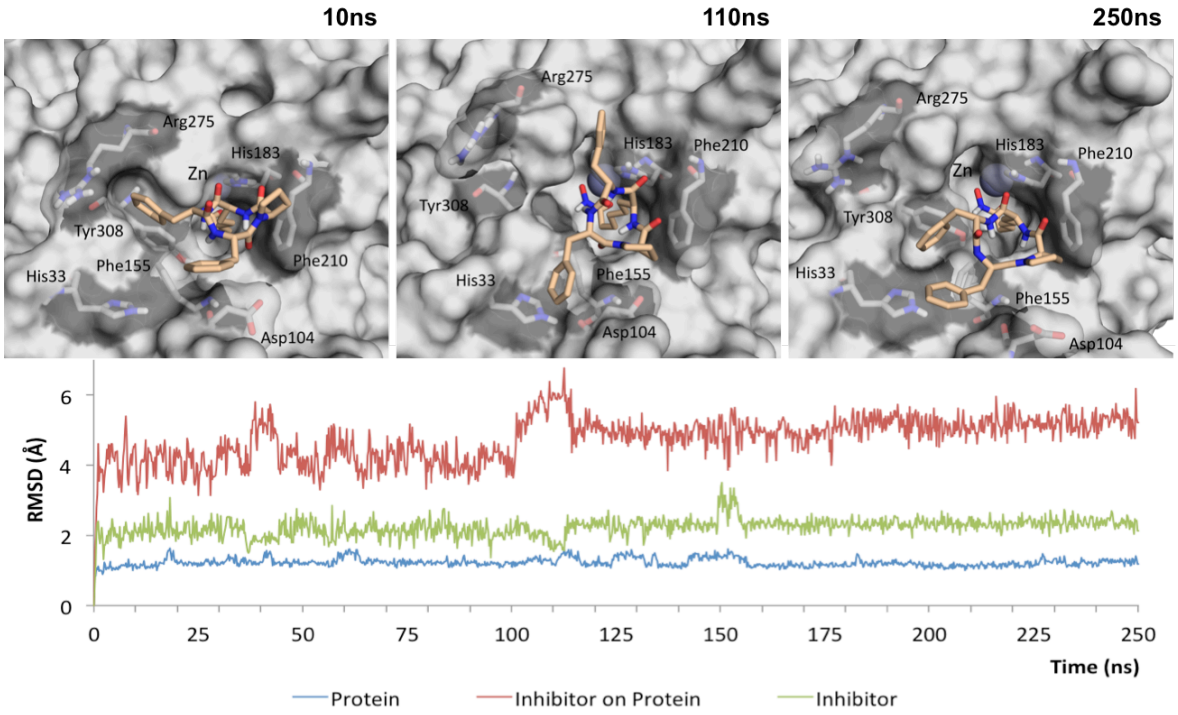


Figure 29 MD simulation snapshots of TpxB^{Asuha} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

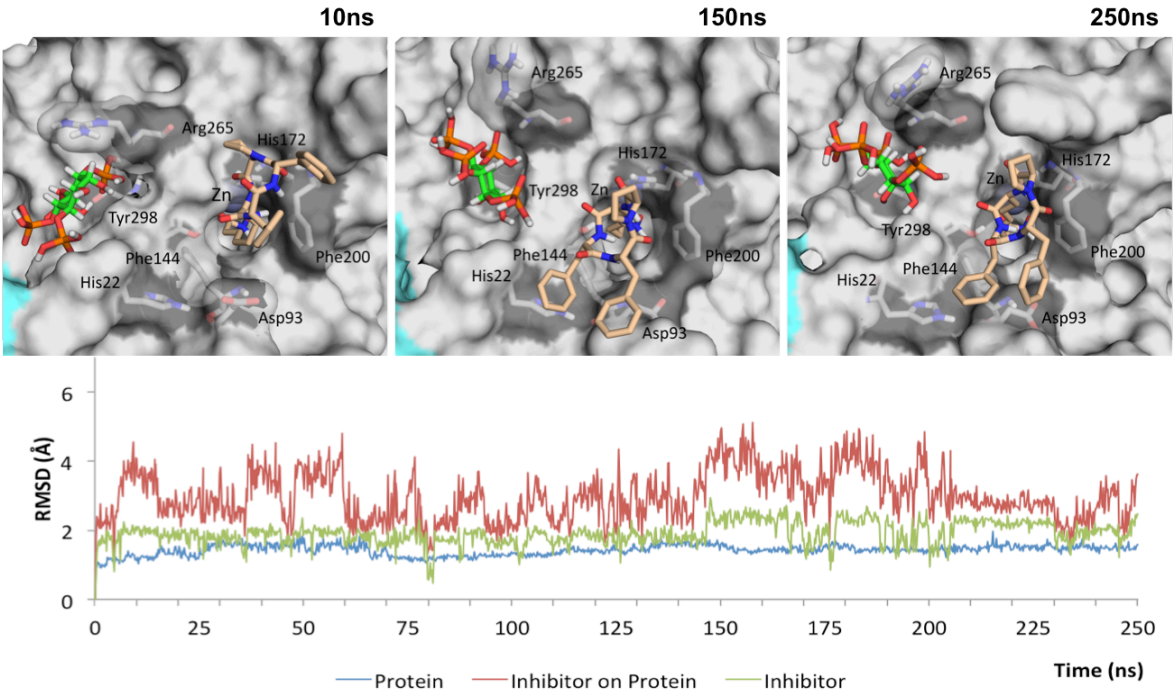


Figure 30 MD simulation snapshots of TpxB^{Asuha} in HDAC3 and RMSD variation throughout 250ns of simulation. Inositol phosphate molecule shown as sticks with greeen carbons and co-repressor surface shown in blue.

Cyl-1^{Asuha} position in HDAC2 binding pocket starts very unstable with only hydrophobic interactions with Phe155 and Phe210 keeping the macrocycle in place (Figure 31). 12ns after the simulation started, the macrocycle moves slightly towards Asp104 and bonds to its side chains through a water bridge. Water bridges are also maintained with Phe210, with might prevent to macrocycle from positioning closer to Asp104 and form hydrogen bonds. Similar results were found for Cyl-1^{Asuha} in HDAC3 (Figure 32).

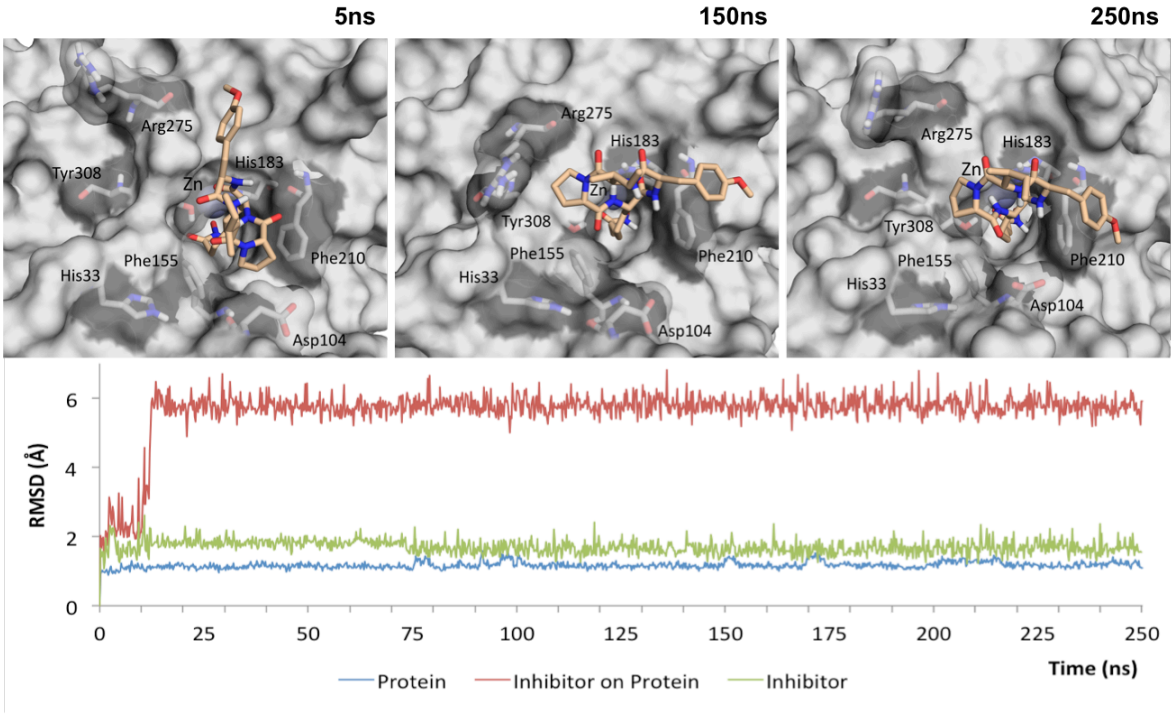


Figure 31 MD simulation snapshots of Cyl-1^{Asuha} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

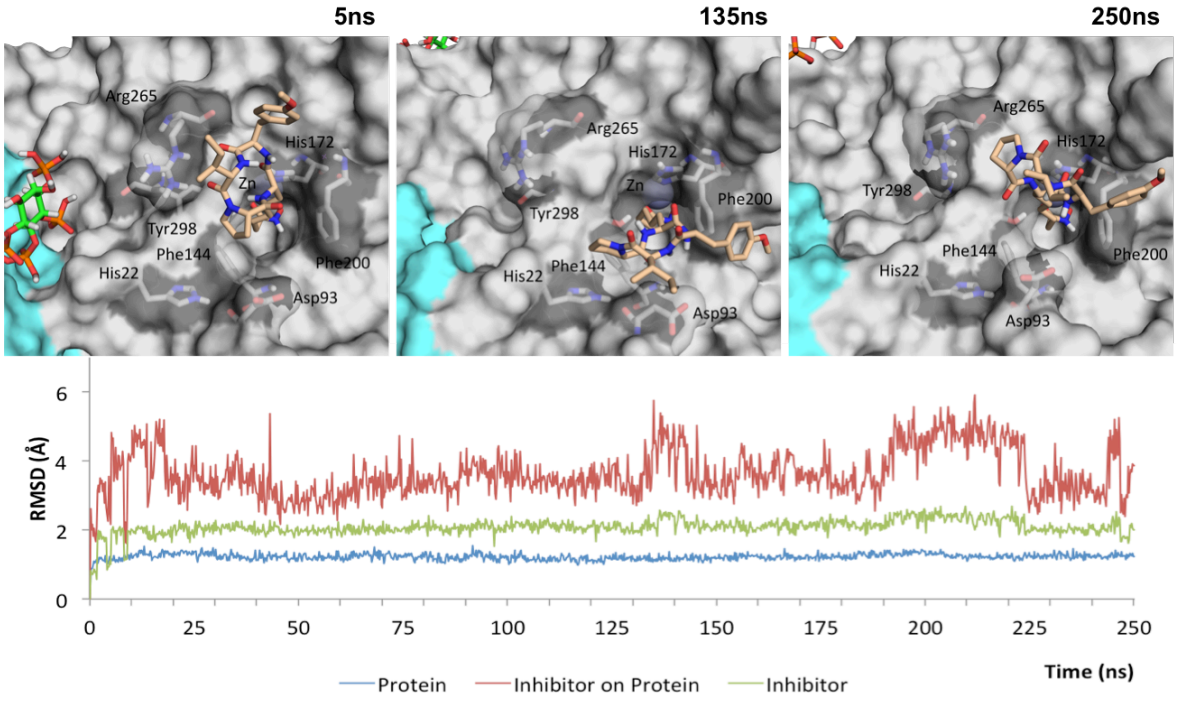


Figure 32 MD simulation snapshots of Cyl-1^{Asuha} in HDAC3 and RMSD variation throughout 250ns of simulation. Details as in **Figure 28**.

Similarly to Cyl-1^{Asuha}, Cyl-2^{Asuha} macrocycle is also positioned away from Asp104/93. During both simulations (Figure 33 and Figure 34), the most frequent interactions are water bridges and π-π stacking with Phe210/200. In HDAC2, 19a occasionally establishes water bridges with Asp104 while, in HDAC3 (Figure 34), these interactions are slightly more frequent. However, the general position of the macrocycle hinders the formations of stronger bonds with Asp104/93.

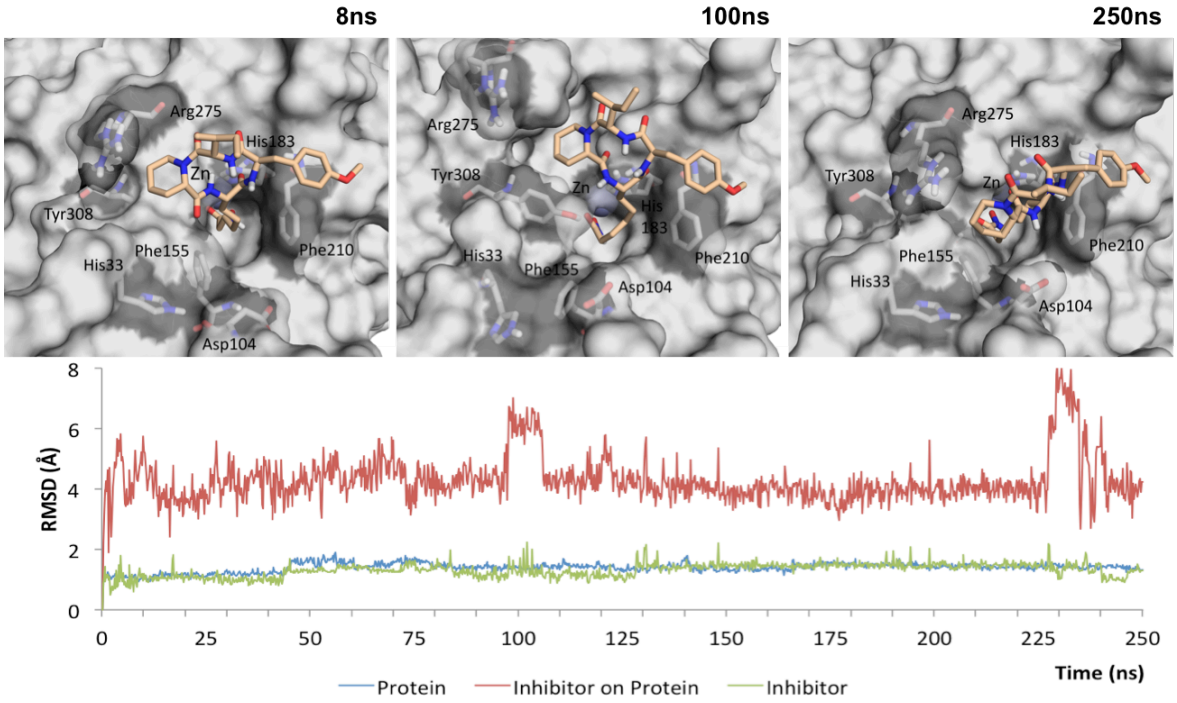


Figure 33 MD simulation snapshots of Cyl-2^{Asuha} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

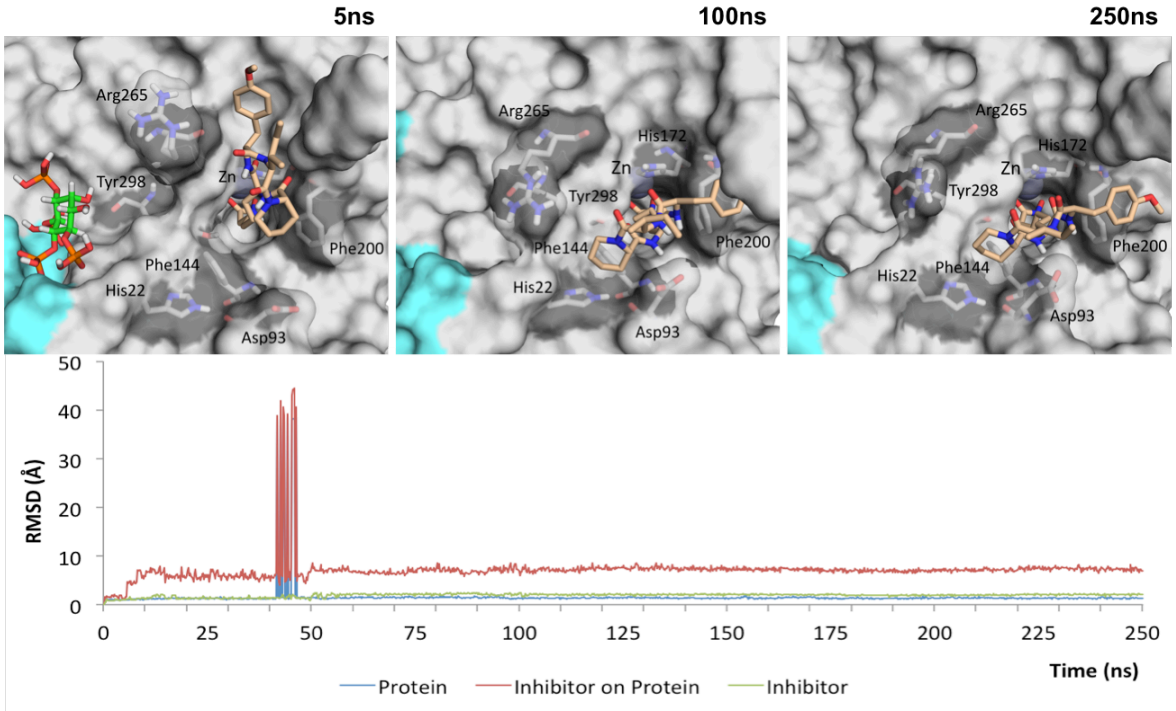


Figure 34 MD simulation snapshots of Cyl-2^{Asuha} in HDAC3 and RMSD variation throughout 250ns of simulation. Details as in **Figure 28**.

According to RMSD analysis apicidin and apicidin A are the most stable inhibitors in both HDAC2 and 3, establishing strong hydrogen bonds with the carboxylate side chain of Asp104/93 and hydrophobic interactions with Pro34/33 and His33/22 aromatic side chain throughout the 250ns of simulation time (Figure 35, Figure 36, Figure 37 and Figure 38). In the case of HDAC3, apicidin A also interacts with the inositol phosphate molecule *via* hydrogen bond and water bridges, during the last 100ns of simulation time (Figure 38).

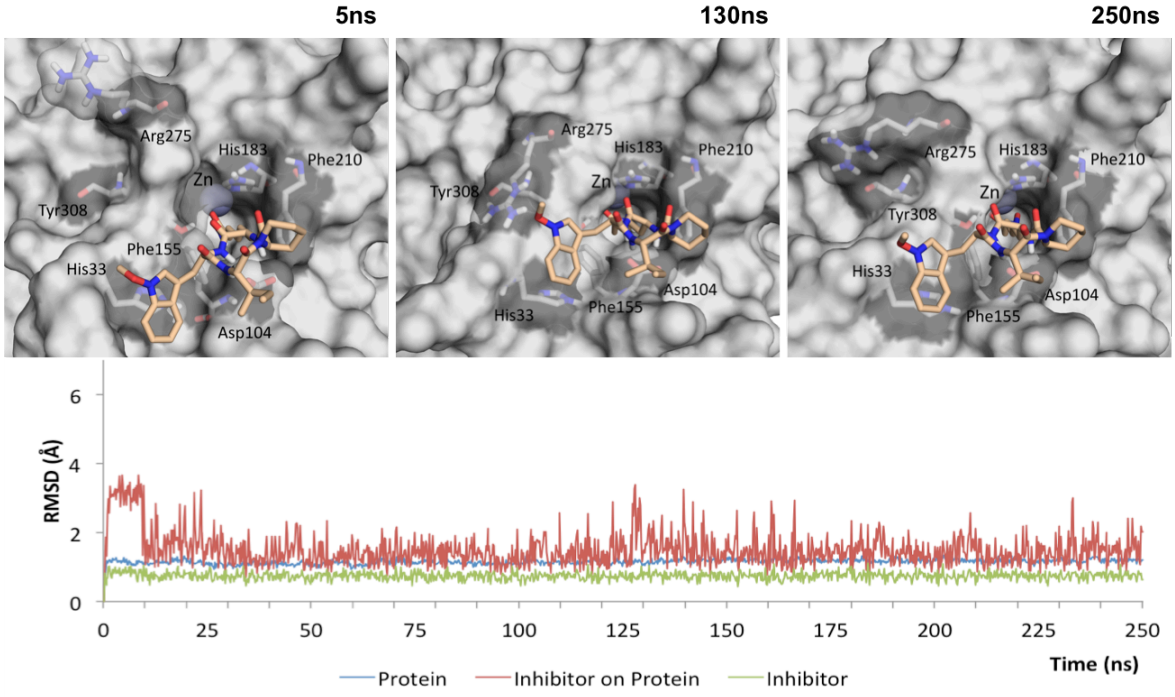


Figure 35 MD simulation snapshots of Api^{Asuha} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

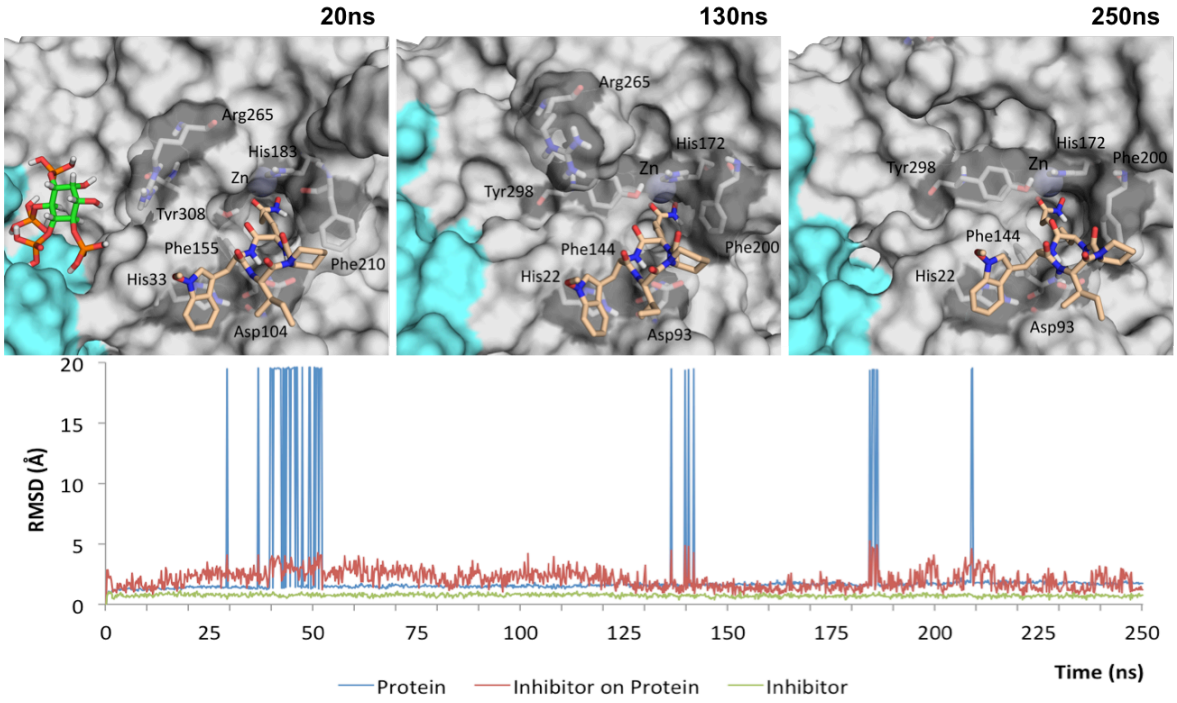


Figure 36 MD simulation snapshots of Api^{Asuha} in HDAC3 and RMSD variation throughout 250ns of simulation. Details as in **Figure 28**.

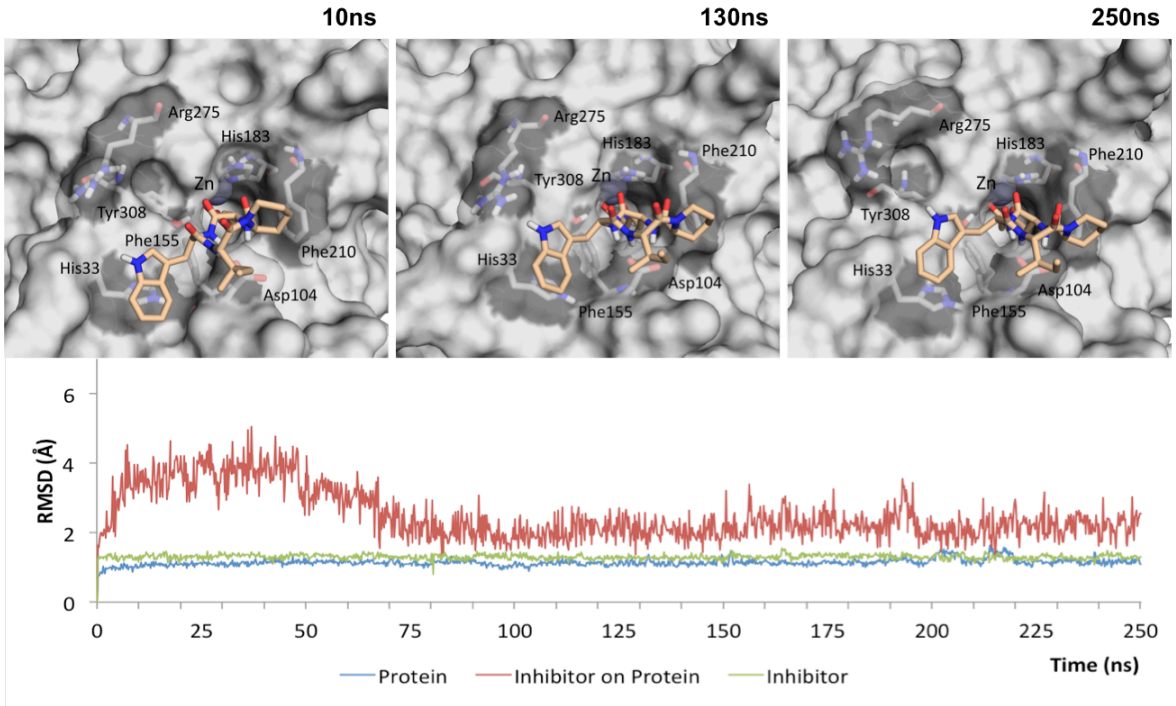


Figure 37 MD simulation snapshots of ApiA^{Asuha} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

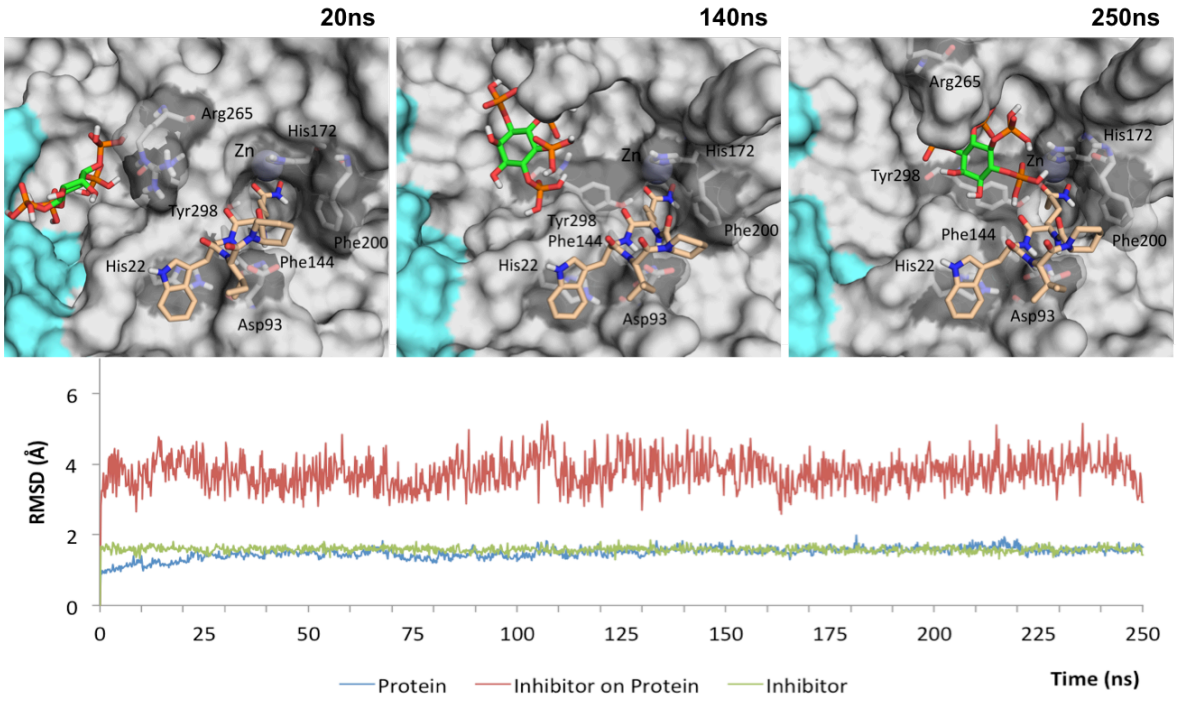


Figure 38 MD simulation snapshots of ApiA^{Asuha} in HDAC3 and RMSD variation throughout 250ns of simulation. Details as **in Figure 28.**

In all simulations performed on hydroxamic acid analogue complexes, the inhibitor binding mode is maintained and the ligands remain bonded throughout the simulation time. In general, interactions between the ZBG and the residues inside the binding pocket stay the same in all six analogues. These interactions include hydrogen bonds with Tyr308/298, His145/134 and His146/135. On the other hand, protein-ligand interactions on the enzyme surface vary significantly depending on the inhibitor: the most stable ones (Api and ApiA) form strong hydrogen bonds with Asp104/93, while the others are positioned differently and cannot efficiently interact with Asp104/93.

Contrarily to the hydroxamic acid analogues, norvaline analogues seem to have a much weaker binding affinity towards HDAC2 and HDAC3. The binding mode predicted during the docking experiments is maintained while the inhibitors block the pocket. Similar poses can be observed for each analogue on the two different enzymes. However, from the systems here mentioned, only Cyl-1^{Nva}, Api^{Nva} and ApiA^{Nva} stayed bounded to HDAC2 for the entirety of the simulation time.

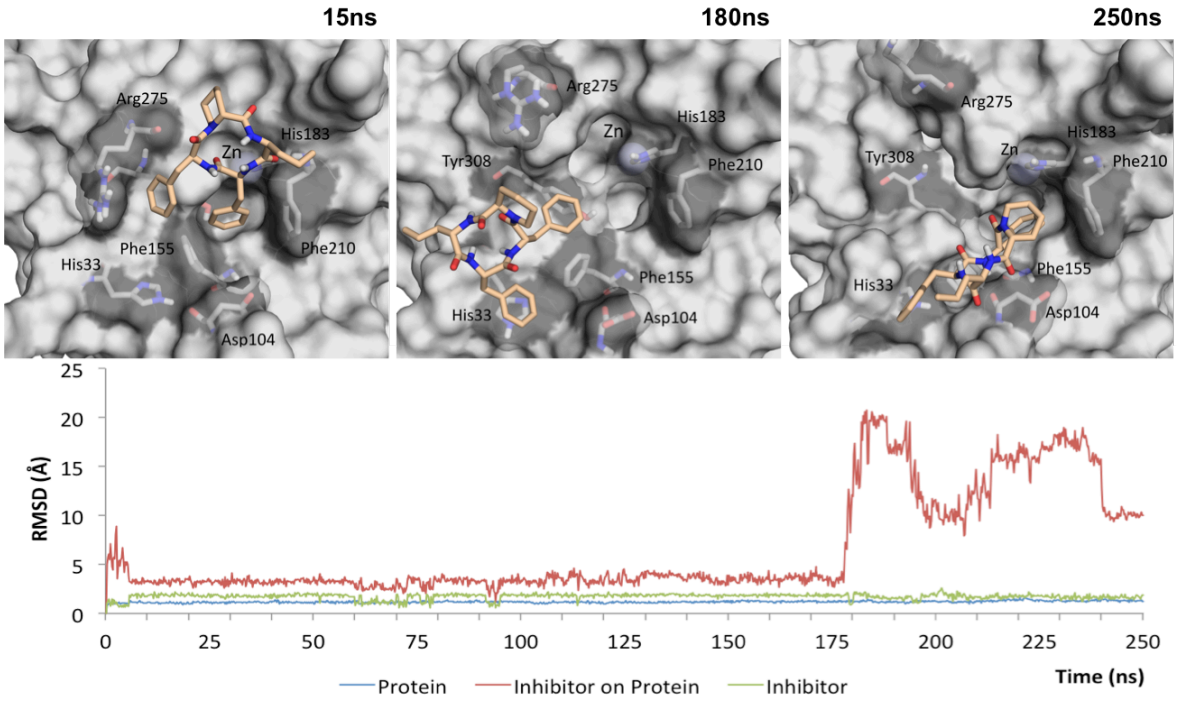


Figure 39 MD simulation snapshots of TpxA^{Nva} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

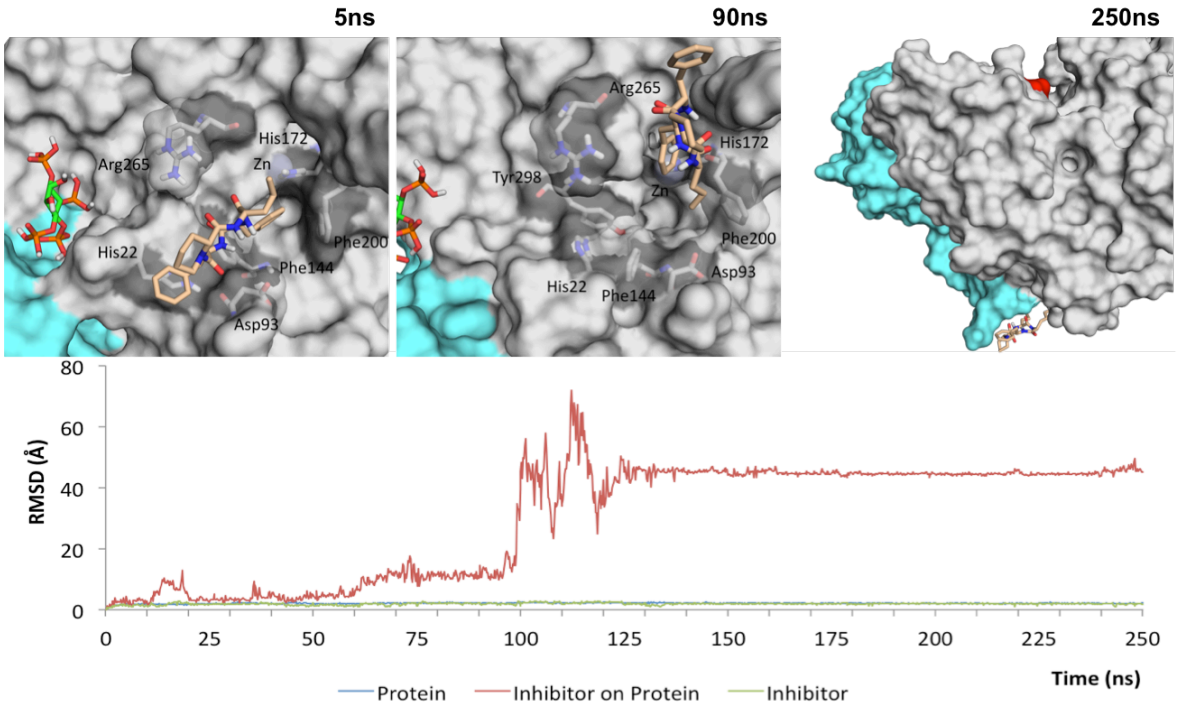


Figure 40 MD simulation snapshots of TpxA^{Nva} in HDAC3 and RMSD variation throughout 250ns of simulation. Details as in **Figure 28**.

While bond in the HDAC2 binding site, TpxA^{Nva} mostly interacts with Phe155, His183, Phe210 and Leu276 (Figure 39). At 179ns the ligand leaves the pocket and begins to interact with Tyr308, Arg311 and Tyr341, returning to partially block the pocket again at 241ns, this time establishing water bridges with Asp104 and hydrophobic interactions with His33. Although TpxA^{Nva} only stays in the pocket for the first 62ns during simulation of HDAC3, similar interactions are established with Phe144 and Phe200 (Figure 40). After this, the inhibitor moves away from the pocket and goes into solution until the 127ns,time at which goes back to interact with HDAC3 at a completely different site (Met45, Val47 and Met416).

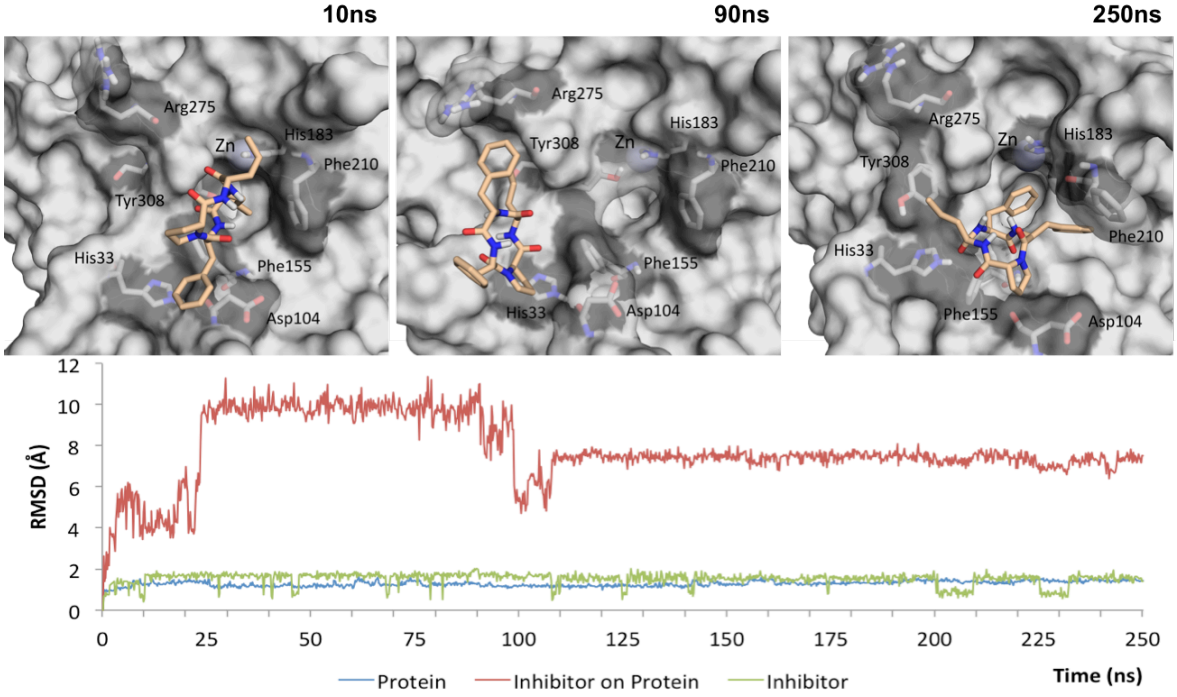


Figure 41 MD simulation snapshots of TpxB^{Nva} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

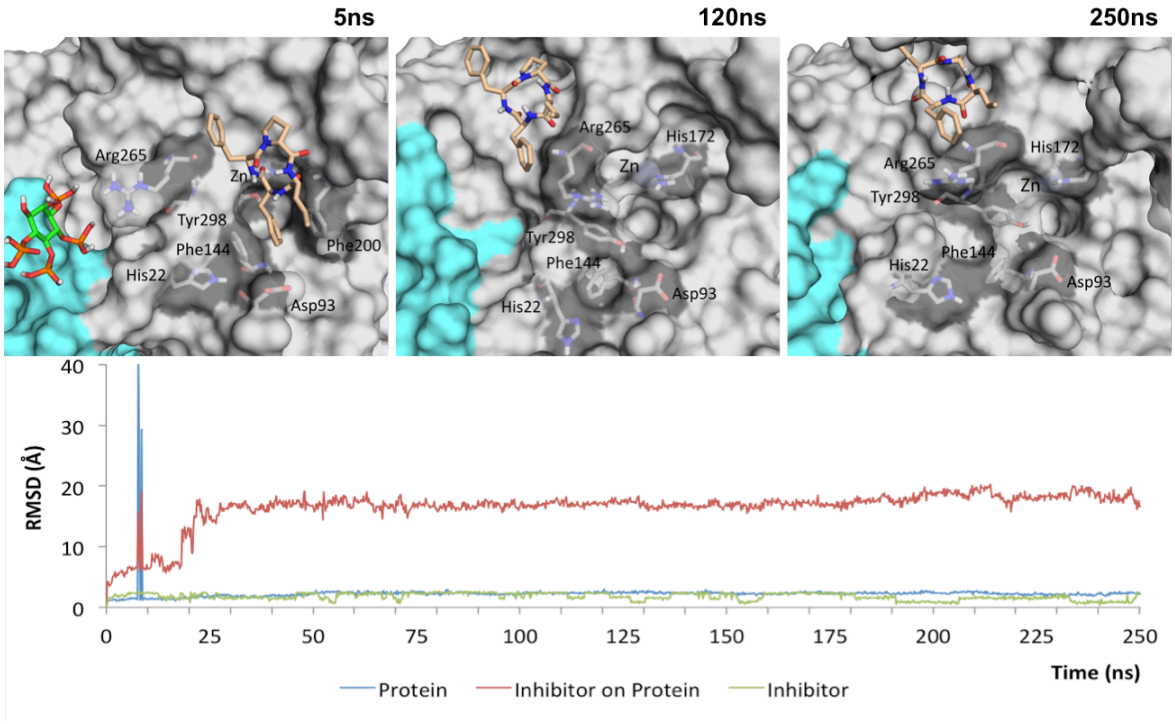


Figure 42 MD simulation snapshots of TpxB^{Nva} in HDAC3 and RMSD variation throughout 250ns of simulation. Details as in **Figure 28**.

In both enzymes, TpxB^{Nva} starts bound in the binding site interacting with Asp104/93, Phe155/144 and Phe210/200 (Figure 41 and Figure 42). At 24ns (19ns in HDAC3) the inhibitor moves away from the pocket entrance, leaving it open to accommodate other ligands. In the case of HDAC3, TpxB^{Nva} remains bonded in the same place until the end of the simulation, establishing hydrogen bonds with Cys263 and Arg265, and a water bridge with Arg301 (Figure 42a). On the other hand, in the HDAC2 simulation, after the 99ns mark the inhibitor goes back to interact with Asp104, His183, Phe155 and Phe210, occupying the pocket and preventing binding of other molecules (Figure 41a).

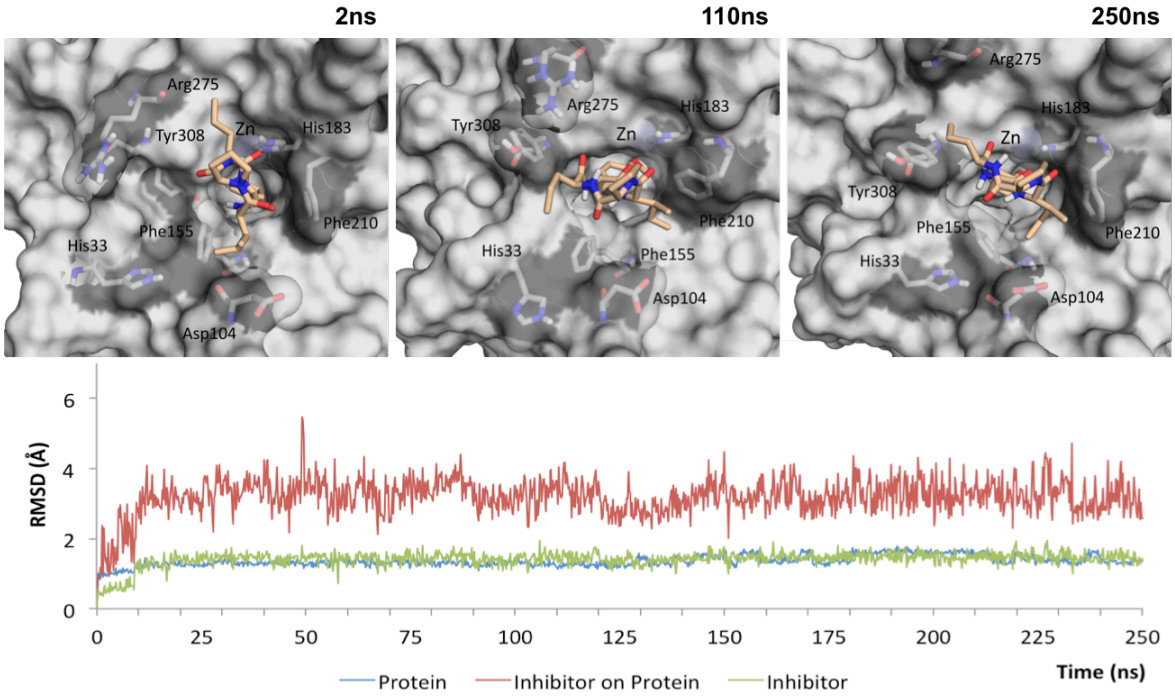


Figure 43 MD simulation snapshots of Cyl-1^{Nva} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

As one of the few exceptions, Cyl-1^{Nva} stays bound in the HDAC2 pocket, blocking it from other ligands, throughout the 250ns of simulation time (Figure 43). This pose is maintained by strong ionic interactions between the methoxy group in the Tyr side chain and the catalytic Zn^{2+} ion, as well as π-π stacking and water bridges with His183. In HDAC3 (Figure 44), and although Cyl-1^{Nva} starts in a similar pose to HDAC2, the interactions established with Asp93 and Phe144 are not strong enough to keep the ligand bonded. At 173ns the pocket is free and Cyl-1^{Nva} barely interacts with HDAC3, being mostly in the solvent.

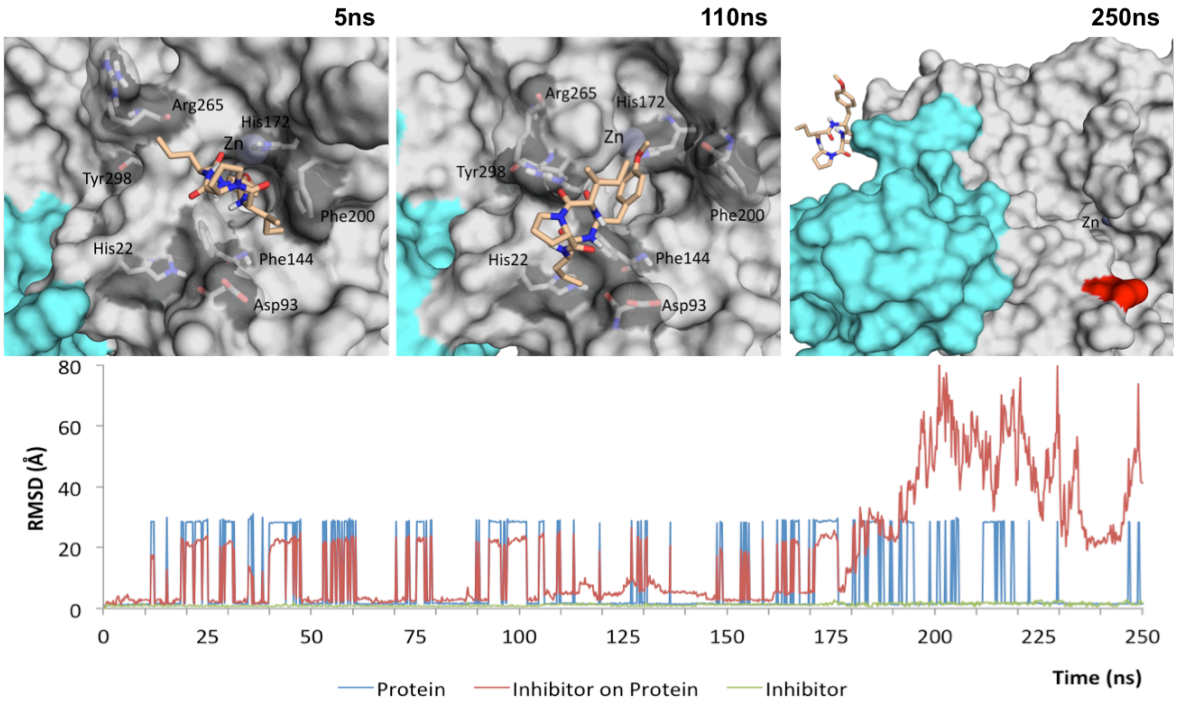


Figure 44 MD simulation snapshots of Cyl-1^{Nva} in HDAC3 and RMSD variation throughout 250ns of simulation. Position of ASP93 shown as red surface. Details as in **Figure 28**.

Although Cyl-2^{Nva} starts by establishing similar interactions to its hydroxamic acid counterpart in HDAC2 (hydrogen bonds to Asp104, π-π stacking with Phe155 and Phe210, and water bridges with Arg275 and Tyr308), after 72ns Cyl-2-Nva leaves the binding pocket, which is now free for binding of other molecules (Figure 45). Cyl-2^{Nva} goes into solution at 118ns and returns to interact with Glu340 at 134ns. The inhibitor stays bound to Ile310 and Arg311 from 164ns to the end of the simulation.

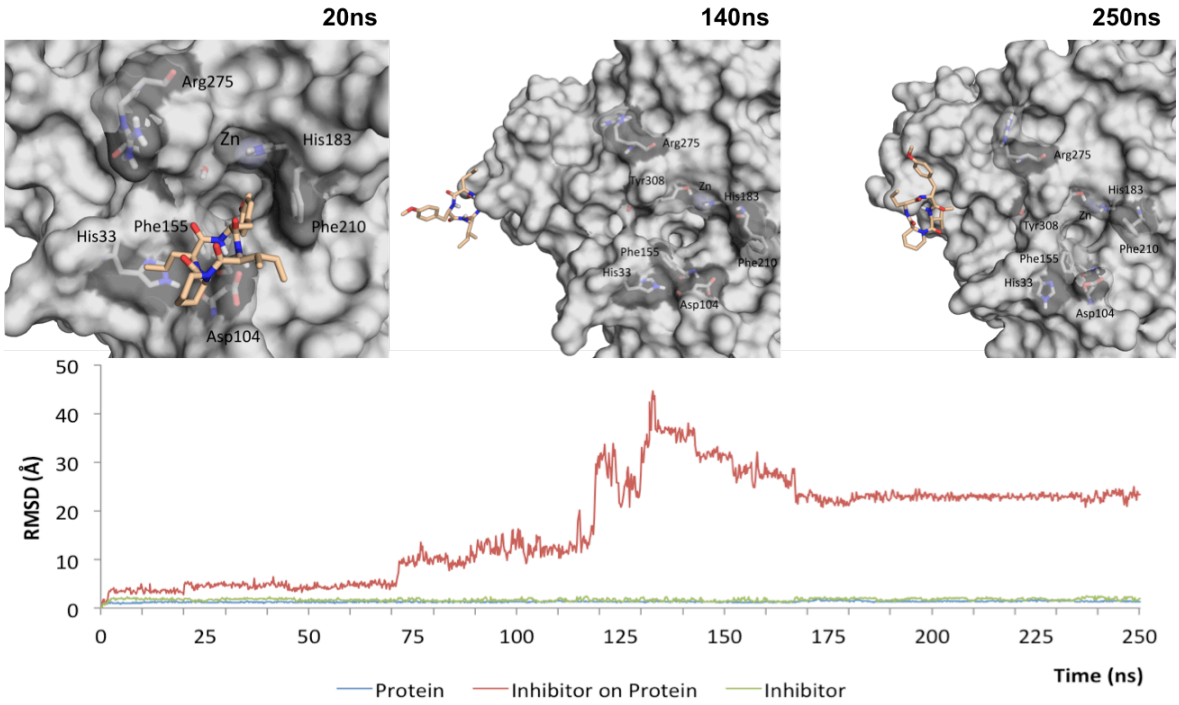


Figure 45 MD simulation snapshots of Cyl-2^{Nva} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

In the case of HDAC3, the binding affinity seems to improve slightly. Cyl-2-Nva stays bounded for 213ns, interacting with the Zn^{2+} ion and Asp93 through ionic interaction and hydrogen bonds, respectively (Figure 46). Water bridges with Tyr298 and π-π stacking with Phe144 keep the Tyr(Me) side chain in the pocket until the 72ns, after which the side chain moves towards the enzyme surface, blocking the pocket only partially. Despite of this, only after 195ns of simulation time, Cyl-2^{Nva} moves completely away from the pocket.

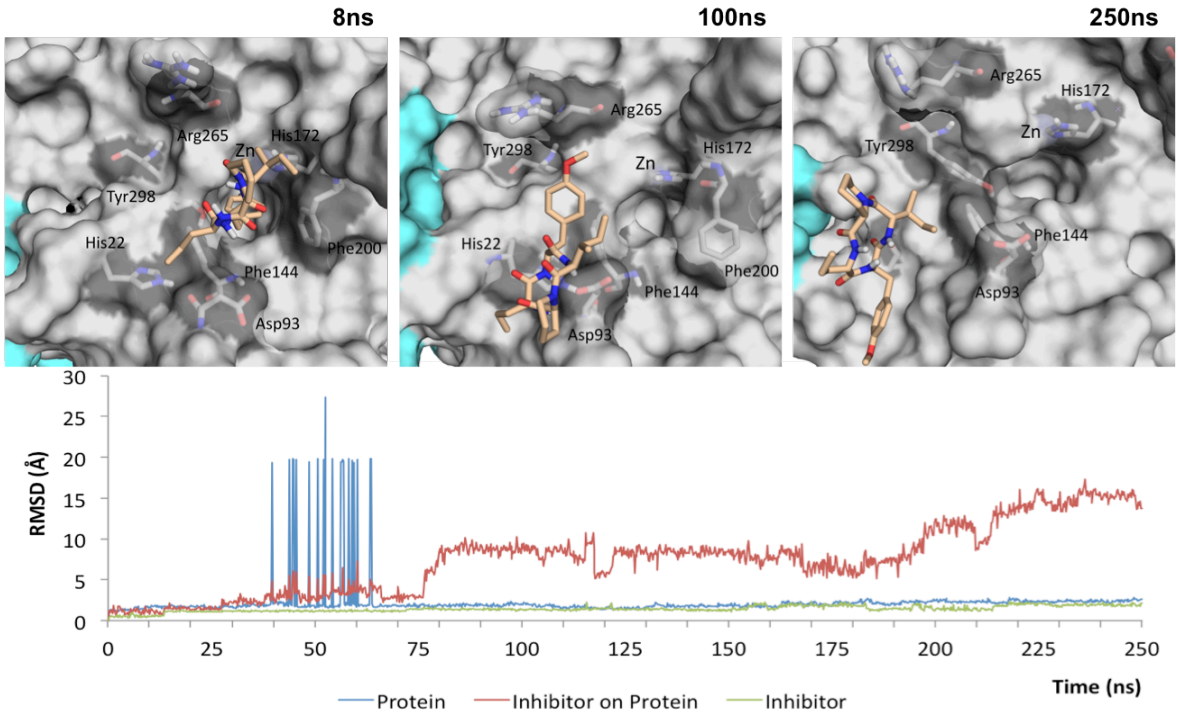


Figure 46 MD simulation snapshots of Cyl-2^{Nva} in HDAC3 and RMSD variation throughout 250ns of simulation. Position of ASP93 shown as red surface. Details as in **Figure 28**.

Regarding Api^{Nva}, interactions with the two enzymes are very similar. Strong hydrogen bonds with Phe210/200 and His183/172 keep the ligand at the entrance of the binding pocket, blocking it (Figure 47 and Figure 48). The water bridge to either Asp104 or Tyr198 induces a change in the orientation of the macrocycle and, in the case of HDAC3, destabilises the inhibitor and causes Api^{Nva} to leave the HDAC3 pocket after 219ns of simulation (Figure 48).

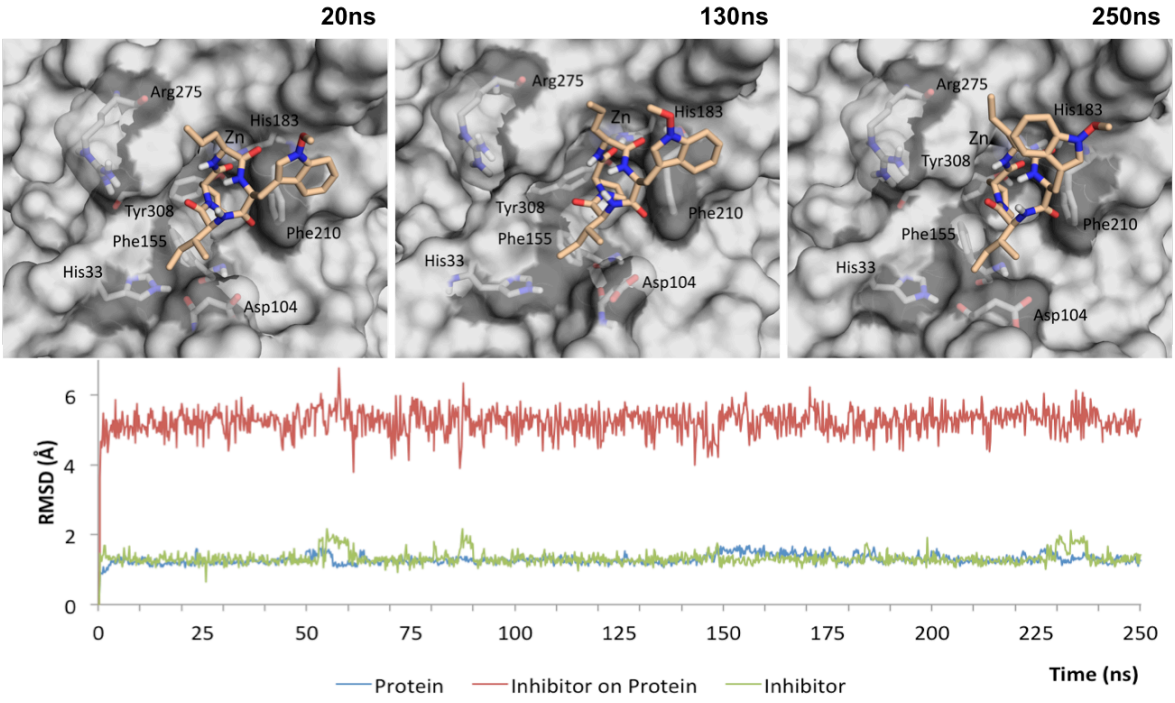


Figure 47 MD simulation snapshots of Api^{Nva} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

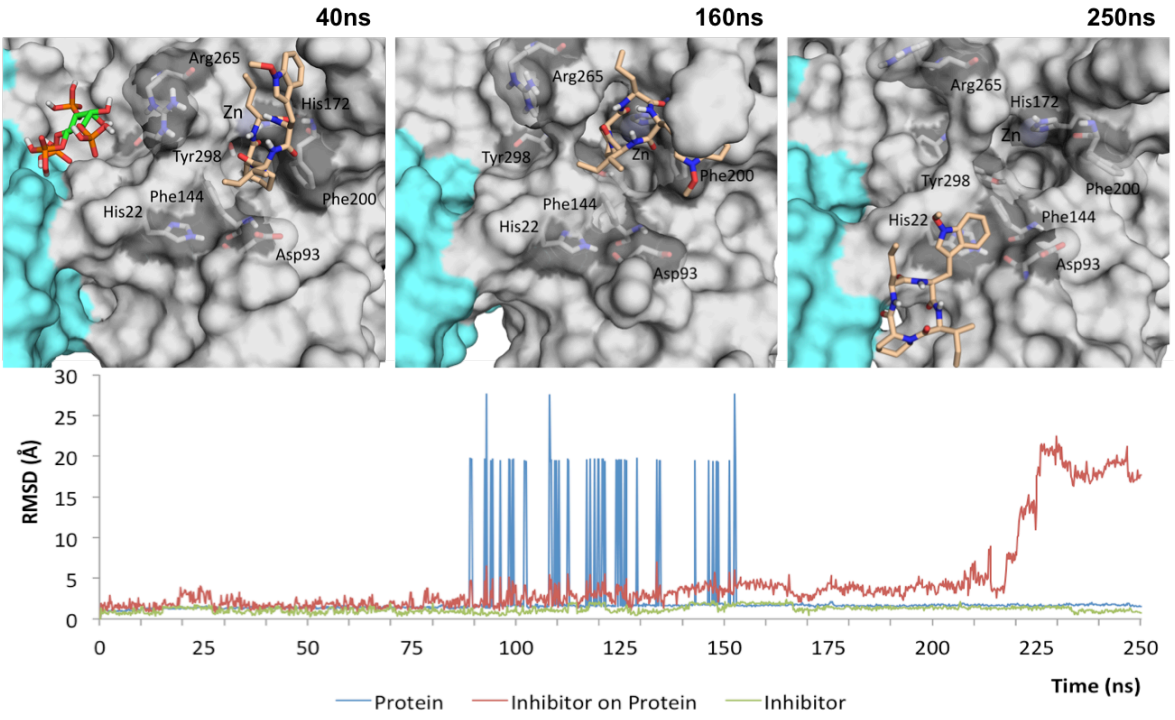


Figure 48 MD simulation snapshots of Api^{Nva} in HDAC3 and RMSD variation throughout 250ns of simulation. Position of ASP93 shown as red surface. Details as in **Figure 28**.

ApiA^{Nva} assumes an identical pose to Api^{Nva} in HDAC2, interacting with Phe155, His184 and Phe210 *via* hydrogen bonds and water bridges, throughout the 250ns of simulation (Figure 49). Water bridges with Asp104 and Leu276 induce rotation of the cap group without tampering with the blocking of the pocket. With HDAC3, the results are surprisingly different. ApiA^{Nva} starts in the same pose as in HDAC2, hydrogen bonding with Phe200. However, after the first 11ns, a change in the inhibitor conformation "pushes" the loop with the Phe200 residue away from the binding pocket and, with the movement of ApiA^{Nva} upwards, the pocket is left completely open to other molecules (Figure 50). After 50ns, ApiA^{Nva} starts interacting with His172 and goes back to partially blocking the pocket. This dynamics repeats itself for the rest of the simulation: His172 interactions favour blocking of the pocket, and interactions with Asn197 and Lys194 "pull" ApiA^{Nva} away from the pocket.

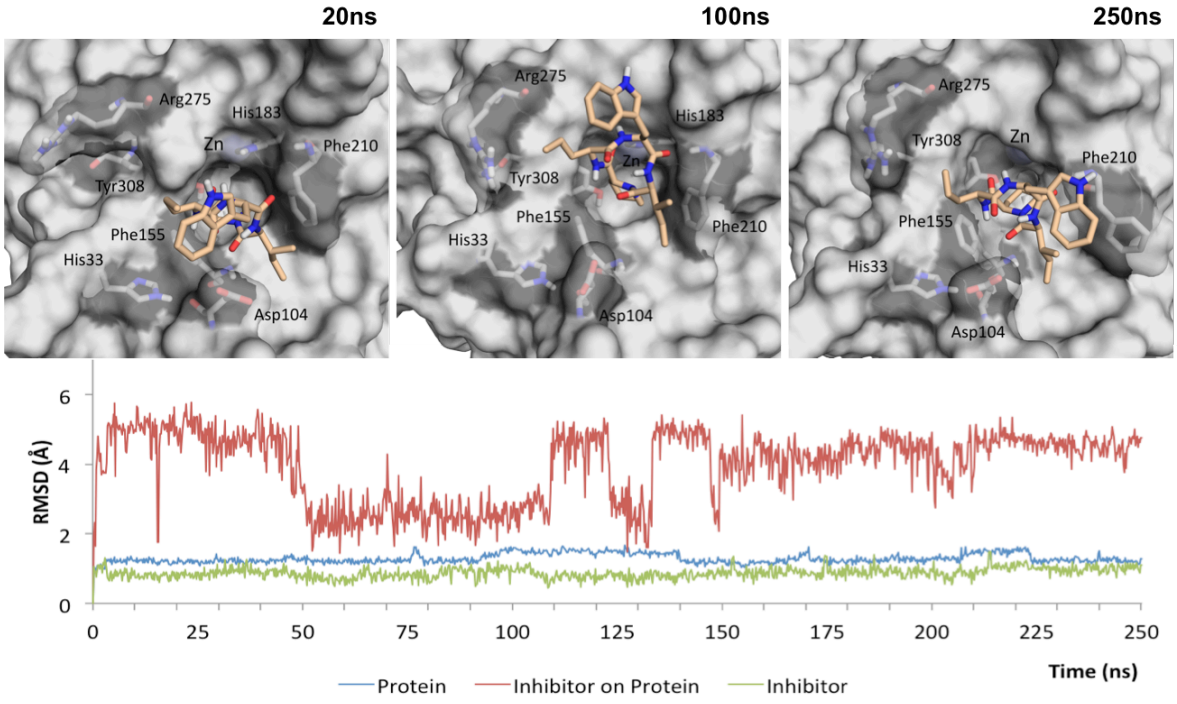


Figure 49 MD simulation snapshots of ApiA^{Nva} in HDAC2 and RMSD variation throughout 250ns of simulation. Details as in **Figure 27**.

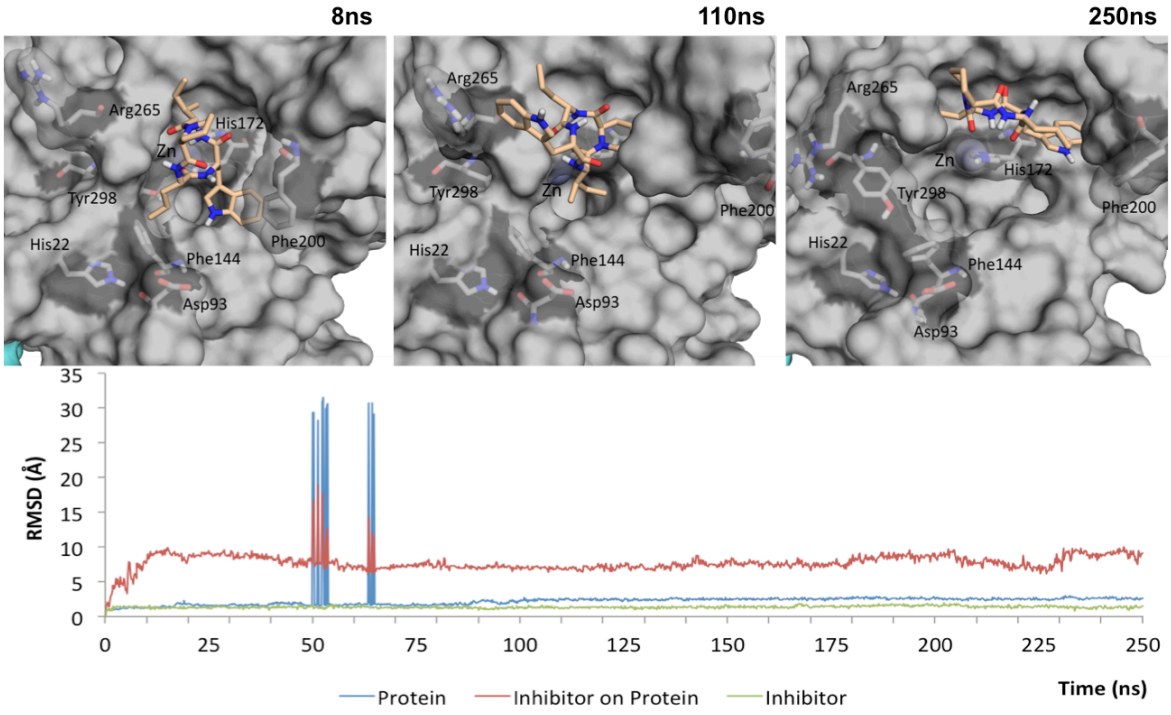


Figure 50 MD simulation snapshots of ApiA^{Nva} in HDAC3 and RMSD variation throughout 250ns of simulation. Position of ASP93 shown as red surface. Details as in **Figure 28**.

Due to the lack of a ZBG, norvaline analogues became much less potent than their hydroxamic acid counterparts. This resulted in a significant loss of binding affinity and production of highly unstable poses during the MD simulations. Apart from Cyl-1^{Nva}, Api^{Nva} and ApiA^{Nva} in HDAC2, all other complexes were unable to stay bound throughout the entire simulation time.

Another finding obtained from trajectory analysis showed abnormally increased RMSD values for HDAC3 protein backbone and ligand position. To understand the origin of the abnormal root-mean-square deviation (RMSD) values, the root-mean-square fluctuation (RMSF) per residue was calculated. As expected, results showed the most flexible parts of the protein were the N- and C-termini as well as the co-factor binding loop. Unfortunately, these data could not explain the previous findings. In fact, according to the evidence in Figure 51, the positional fluctuations revealed by the RMSF per residue were limited to the first 40ns of simulation, which differs from the high RMSD values identified throughout the MD simulations with HDAC3. Additional manual analysis of frames with the unintelligible increase in RMSD revealed it to be an artefact. Due to the protein dynamics, in some frames a few residues from the catalytic domain, together with the co-factor, are viewed in another cell

unit. Periodic boundary conditions (PBC) unaware tools, like RMSD calculation with maestro, cannot re-centre the trajectory before performing the calculation, resulting in increasingly higher RMSD values. However, this artefact does not interfere with the actual trajectory and we can, therefore, consider the systems equilibrated and trust the analysis of protein-ligand interactions and dynamics.

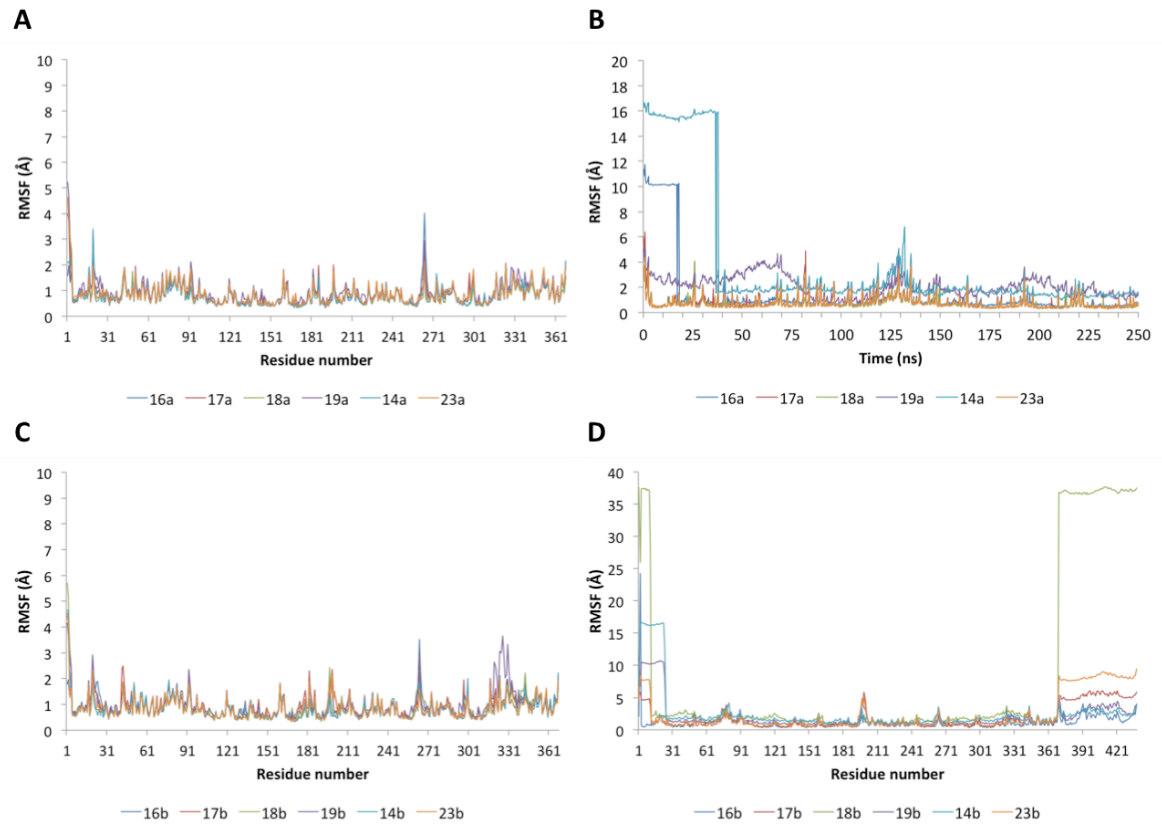


Figure 51 Protein RMSF calculated for heavy atoms of each residue. Left column shows the results for HDAC2 while the right column represents HDAC3. **A** and **B** correspond to hydroxamic acid analogues and **C** and **D** norvaline analogues. Statistic values on tables are given in Å and were calculated for the 250ns of simulation time.

2.4 Conclusion

In this part of the project we investigated the binding mode and affinity of a collection of known macrocylic HDAC inhibitors and their analogues with a truncated ZBG.

The results from the conformational search revelead that, in general, Cyl-1 and Cyl-2 assume a TCTT conformation whereas trapoxins and apicidins preferentially exhibit TTTT conformations. However, the tertiary amide proved to be highly flexible, visiting both *trans* and *cis* conformations in all of the SA experiments. The preferred conformations proved to be dependent not only on the solvent but also on the presence or lack of a ZBG.

Docking studies showed that, in agreement with a previous study on structurally similar compounds[216], hydroxamate analogues mimic the substrate's binding mode (competitive inhibition) and, as expected, yield the best docking scores. On the other hand, norvaline analogues project the aromatic side chains towards the active site channel, which could explain the non-competitive inhibition result from experimental data. Findings suggest Asuha trapoxins and apicidins as the highest affinity inhibitors, while Cyl-1 and Cyl-2 show the lowest scores amongst the norvaline analogues. Careful analysis of the docking poses and conserved interactions across all of them also revealed that more interactions do not necessarily mean potentially higher affinity inhibitors. The most favourable poses, within the hydroxamic acid analogues, seem to happen when the inhibitor interacts with both Asp104/93 (hydrogen bonds) and His33/22 (π-π stacking). These interactions stabilize the macrocycle, on the enzyme surface, and allow for a better positioning of the linker and ZBG, and better interaction with the catalytic Zn^{2+}. For norvaline analogues, interactions with Asp104/93 and His33/22 do not favour high affinity binding. Interaction with these residues pulls the inhibitors away from the pocket entrance and, because of the shortening of the linker and the lack of a ZBG, the inhibitor can no longer block the pocket completely. The lack of a ZBG also impairs the stability of the macrocycle on the enzyme surface.

After each one of the best docking poses was submitted to MD simulations, all the hydroxamic acid analogues stayed bound to HDAC2 and 3, chelating to the Zn^{2+}, throughout the 250ns of simulation time. Similarly to azumamides[216], strength and frequency of the interactions between the macrocycle backbone amide groups and the carboxylate side chain of Asp104/93, as well as interaction between the aromatic side chain of the third amino acid and His33/22 side chain, seem to correlate with higher affinity of the inhibitor like apicidin and apicidin A. On the other hand, norvaline analogues result in unstable simulations and most of them end up leaving the pocket. Interestingly, Cyl-1^{Nva} and Cyl-2^{Nva} are the only ones able to interact directly with the Zn^{2+} ion. Based on the findings, we hypothesize the reason why Cyl-1^{Nva} and Cyl-2^{Nva} previously showed a slightly better docking score than the trapoxins and apicidins in this study, rests on the interaction between the methoxy group in the macrocycle side chain and the Zn^{2+} ion. Although this finding is not supported by current experimental results, further investigation of this kind of motif might provide important insight on HDAC binding preferences. The knowledge of the key interactions established between ligand and protein will enable us to develop new inhibitors with high selectivity to specific subtypes within this class of enzymes.

3. NAD+ and NADH effect on sirtuin activity

Nicotinamide adenine dinucleotide (NAD^+) is a coenzyme found in all living cells that mediates redox reactions and participates in the regulation of NAD^+-dependent metabolic pathways[217]. These pathways include calcium signalling *via* CD38/157 ectoenzymes[218], DNA^+ repair through poly-ADP-ribose polymerases (PARPs)[219], expression of genes involved in the circadian clock homeostasis[220] and protein deacetylation by sirtuins[221].

As previously mentioned, sirtuins are a family of Sir2 homologous enzymes featuring a highly conserved catalytic domain where a molecule of NAD^+ binds to catalyse the deacetylation of histone and non-histone proteins. Amongst the many functions known to sirtuins, these enzymes are believed to regulate the life span of many different organisms by responding to changes in the nutritional cell environment. Many studies have shown that caloric restriction and increased NAD^+/NADH ratio promote life span extension by increasing sirtuin activity. This results in the activation of nuclear transcription profiles that increase metabolic efficiency and decrease the production of ROS by promoting SOD2 and insulin degrading enzyme 2 (IDE2) activity in the mitochondria[105,222]. Similarly, decreases in NAD^+ levels caused by the organism's aging can downregulate sirtuin activity[223,224]. To date, the longevity effects of sirtuins have been demonstrated in yeast, worms, flies and mice[225,226]. Moreover, recent studies have shown that low levels of NAD^+ are accompanied by decrease in sirtuin activity and result in normal aging[227,228]. Interestingly, administration of NAD^+ precursors and intermediates was able to normalise the NAD^+ deficiency and associated pathologies. Due to their key role in the regulation of life span, sirtuins have been suggested as therapeutic targets for several age-related diseases like diabetes, cancer and cardiovascular and neurodegenerative diseases[105]. This discovery prompted the investigation of regulatory molecules that might modulate their activity. Several *in vivo* studies have described NADH as a weak competitive inhibitor for yeast sirtuins however to the best of our knowledge, NADH affinity and inhibitory potency had only been be tested for yeast sirtuins and human SIRT2.[229,230]

This chapter is based on the paper by Madsen *et al*.[231] and reports the study of NADH binding to SIRT1, -3 and -5, providing important insights on the sirtuin substrate specificity to be used in the design of inhibitors with improved selectivity. Because of the increasing number of published papers on sirtuin activities and the variety of results reported, a baseline for substrate affinity of the mammalian sirtuins was established agaisnt different lysine modifications was established. The ability of NADH, ADP-ribose (ADPR) and nicotinamide

(NAM) to inhibit these sirtuins was then studied using fluorescence-based assays and HPLC-MS-based assays. Results showed inhibition of all enzyme and substrate combinations at relatively high concentrations (50mM) of NADH. At lower concentrations, deacetylation by SIRT2 and SIRT3 was only slightly inhibited, while deacylation by SIRT1 and SIRT5 was still highly inhibited. ADPR and NAM were also tested under the same conditions as NADH. ADPR exhibited poor inhibition, at 10mM concentration, for most of the enzyme substrate combinations; NAM exhibited inhibition potencies at the μM level, proving to be much more potent than NADH and ADPR.

In order to determine the mechanism by which NADH inhibits the tested sirtuins, deacetylation by SIRT1 and dessucinylation by SIRT5 were measured at varying concentrations of substrate, NAD^+ and NADH. The Lineweaver-Burk plots showed a non-competitive inhibition mechanism agaisnt both substrate and NAD^+ in SIRT1 and SIRT5. To further elucidate the inhibition mechanism and gain insight into the specific binding of NADH, molecular dynamics simulations were performed on the enzyme complexes. This allows the study of the dynamic contributions to protein ligand binding.

3.1 MD simulations

The protein complexes for each sirtuin were obtain by replacing and/or modifying the crystal structures of SIRT1/NAD^+/EX527 (PDBID: 4I5I), SIRT3/FZN (PDBID: 3GLT) and SIRT5/SLL/NAD^+ (PDBID: 3RIY). The systems were neutralized and solvated with TIP3P water and treated using the default pre-simulation protocol in Desmond after which a production run was carried out for 250ns at 300K using the OPLS_2005 force field[1].

Thorough analysis of these simulations revealed a poor parameterization of the NADH molecules. Indeed, in the reduced form the pyridyl nitrogen of NADH should be sp3 hybridized and the nicotinamide ring should adopt a boat conformation. However, due to the use of this force field, the pyridyl nitrogen in NADH is only about 3 degrees out-of-plane. A conformational search of NADH with a ribose unit shows deviations from planarity of 1-5 degrees. When optimizing the same NADH model using DFT-B3LYP, the pyridyl nitrogen moved about 10 degrees out-of-plane and adopted the expected boat conformation. On November 12th 2015, an improved version of the force-field was released (termed OPLS3), and with this forcefield the energy was much more in line with the DFT results when moving

[1] Desmond Molecular Dynamics System, version 3.4, D. E. Shaw Research, New York, NY, 2013; *Maestro-Desmond Interoperability Tools, version 3.4*, Schrödinger, 2013.

the pyridyl nitrogen out-of-plane. Figure 52 shows the truncated NADH model compound and the relative energies (kJ/mol) as result of N-C out-of-plane angle. Both OPLS3 and DFT have a minimum close to 10 degrees, in agreement with previous studies (two references from referees). Based on this we find the OPLS3 significantly better than OPLS_2005 and all of the MD simulations were re-run with the new force field[232,233]. Two replicates were performed for each system.

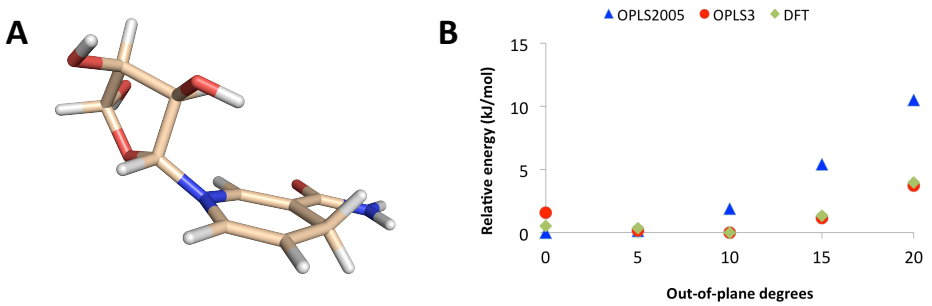


Figure 52 A. NADH model used to test the accuracy of OPLS_2005 and OPLS3 force fields. B. Relative energies (kJ/mol) for the model compound with the pyridyl nitrogen at 0, 5, 10, 15 and 20 degrees out-of-plane.

3.1.1 AC pocket

Previous work[234], demonstrated that the co-substrate NAD+ can bind in two different sites, the AB and the AC pockets. The ADP-ribose moiety is always placed in the A pocket, but the nicotinamide moiety can occupy either the B or the C pocket (Figure 53). The NAD+ binding in the C pocket is essential for deacetylation and if another molecule, such as an inhibitor, binds in this pocket, the reaction does not occur.

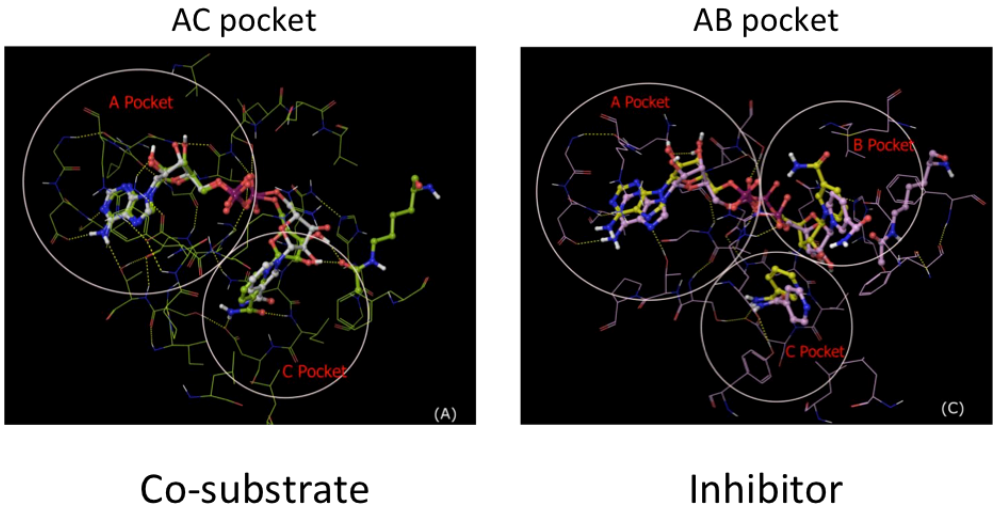


Figure 53 A, B and C binding pockets. In the presence of a peptide substrate, the co-substrate NAD+ binds in the productive conformation in the AC pocket. In the presence of a inhibitor (NAM), NAD+ is forced to bind in the AB pocket. Figure adapted from Guan *et al*[235].

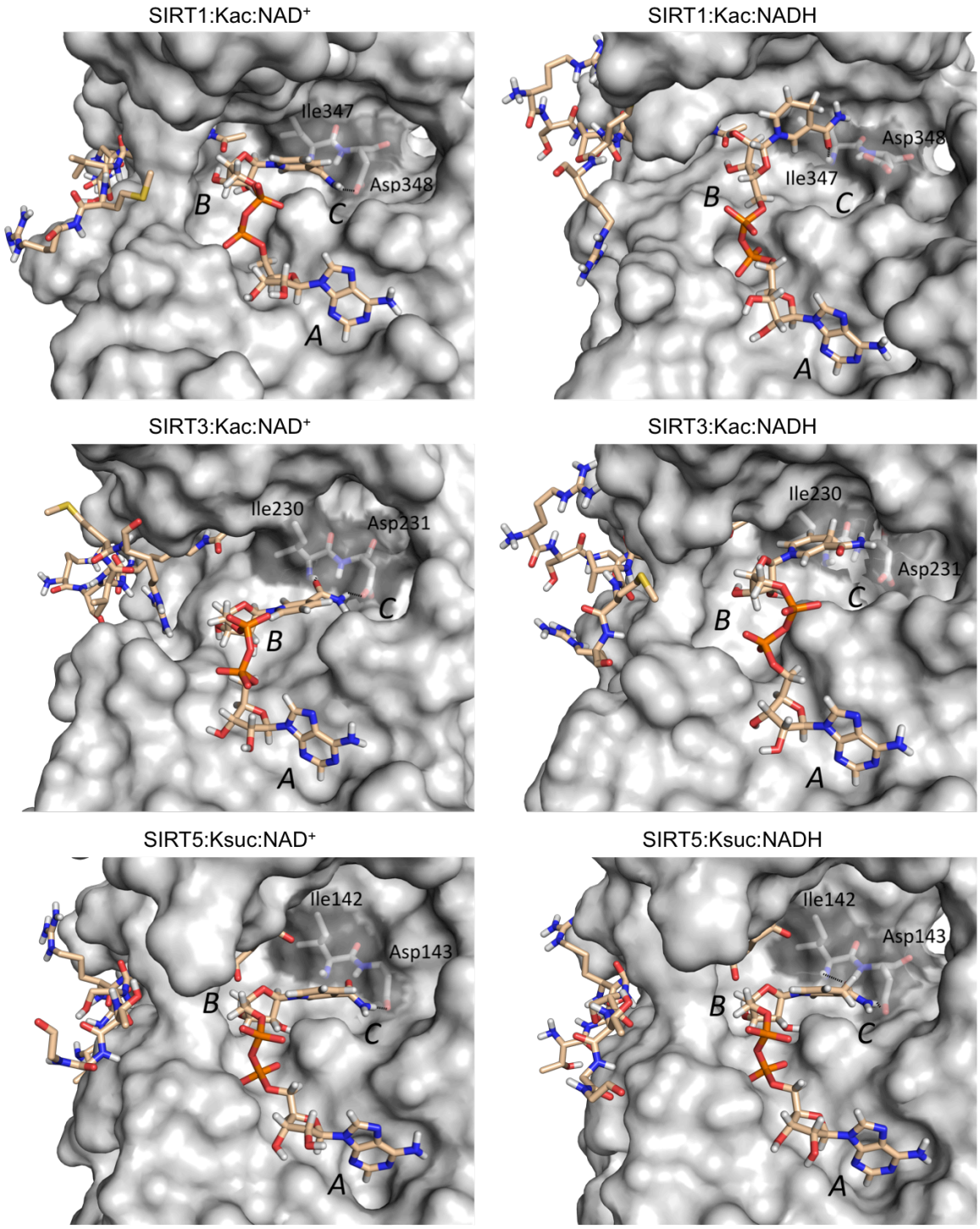


Figure 54 Snapshot of the AC pocket simulations of SIRT1, 3 and 5, after 250ns of simulation time. Protein surface is represented in grey and substrate and co-substrate as sticks. Carbons are colored in wheat, oxygen in red, nitrogen atoms in blue, phospates in orange, sulfur in yellow and hydrogens in white.

In the case of our systems, in the presence of the substrate, NAD+ binds in the AC pocket. This is in agreement with previous findings of an ordered binding[236,237], where peptide substrate binding closes the binding cleft, stabilizes the complex, and allows NAD+ to bind in a productive conformation (i.e. AC pocket). On the other hand, although NADH is initially placed in the AC pocket of SIRT1 and SIRT3, the nicotinamide moiety starts moving away from the C pocket. While in SIRT5, NAD+ and NADH stay bonded in very similar positions in the AC pocket (Figure 54).

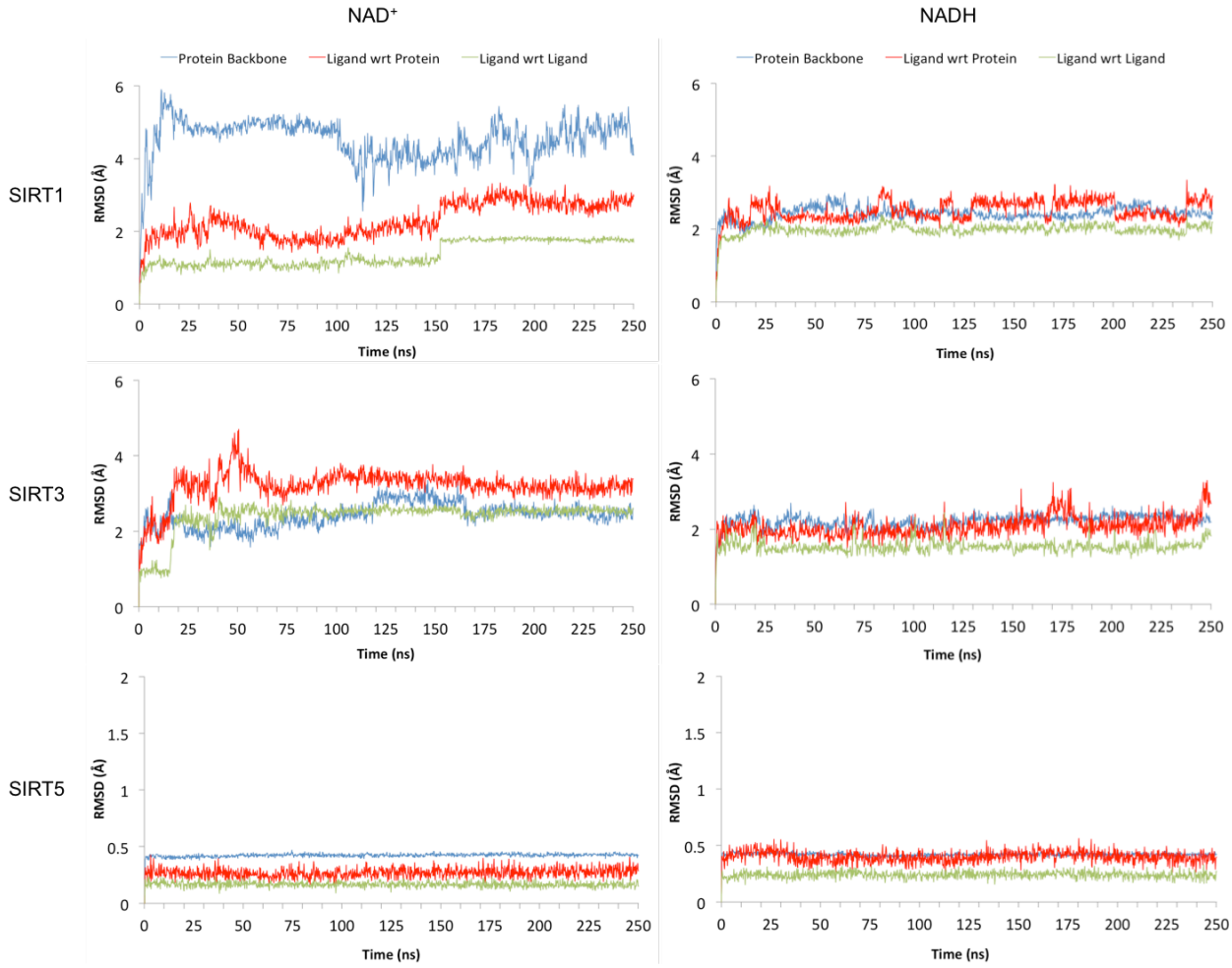


Figure 55 RMSD values, for the protein, over the 250ns of simulation time. The protein backbone is represented as the blue line, the ligand's position with relation to (wrt) the protein is shown in red, and the ligand RMSD is in green.

Table 11 RMSD values (Å) averaged over the last 50ns of simulation time.

	Protein Backbone		Ligand wrt Protein		Ligand wrt Ligand	
	NAD^+	NADH	NAD^+	NADH	NAD^+	NADH
SIRT1	4.68 ± 0.34	2.51 ± 0.13	2.76 ± 0.15	2.50 ± 0.24	1.77 ± 0.04	1.98 ± 0.11
SIRT3	2.47 ± 0.13	2.34 ± 0.09	3.17 ± 0.15	2.24 ± 0.29	2.54 ± 0.08	1.57 ± 0.15
SIRT5	0.43 ± 0.01	0.42 ±0.01	0.28 ± 0.04	0.39 ± 0.04	0.16 ± 0.02	0.23 ± 0.03

Analysis of the RMSD average values during the last 50ns of simulation time, when all systems were equilibrated (Figure 55 and Table 11), showed that NADH tends to be more stable than NAD^+ in SIRT1 and SIRT3 (2.50 ± 0.24 vs 2.76 ± 0.15 and 2.24 ± 0.29 vs 3.17 ± 0.15 for NADH and NAD^+ in SIRT1 and 3, respectively). This indicates that the movement of the NADH nicotinamide moiety away from the C pocket favours the stability of the ligand and of the system in general. This is also in agreement with previous findings supporting NAD^+ destabilization is necessary to facilitate the nicotinamide bond cleavage and the reaction[238]. On the other hand, SIRT5 seems to be higly stable with both ligands, showing RMSD values lower than 1Å. These findings, together with the NADH position during the simulations, suggest a competitive inhibition mechanism for NAD^+ in SIRT5 and, surprisingly, a non-competitive inhibition mechanism in SIRT1 and SIRT3.

The different binding affinities of SIRT1, 3 and 5, for NAD^+ and NADH, were also assessed through the analysis of the hydrogen bonds established between protein and dinucleotide (Figure 56).

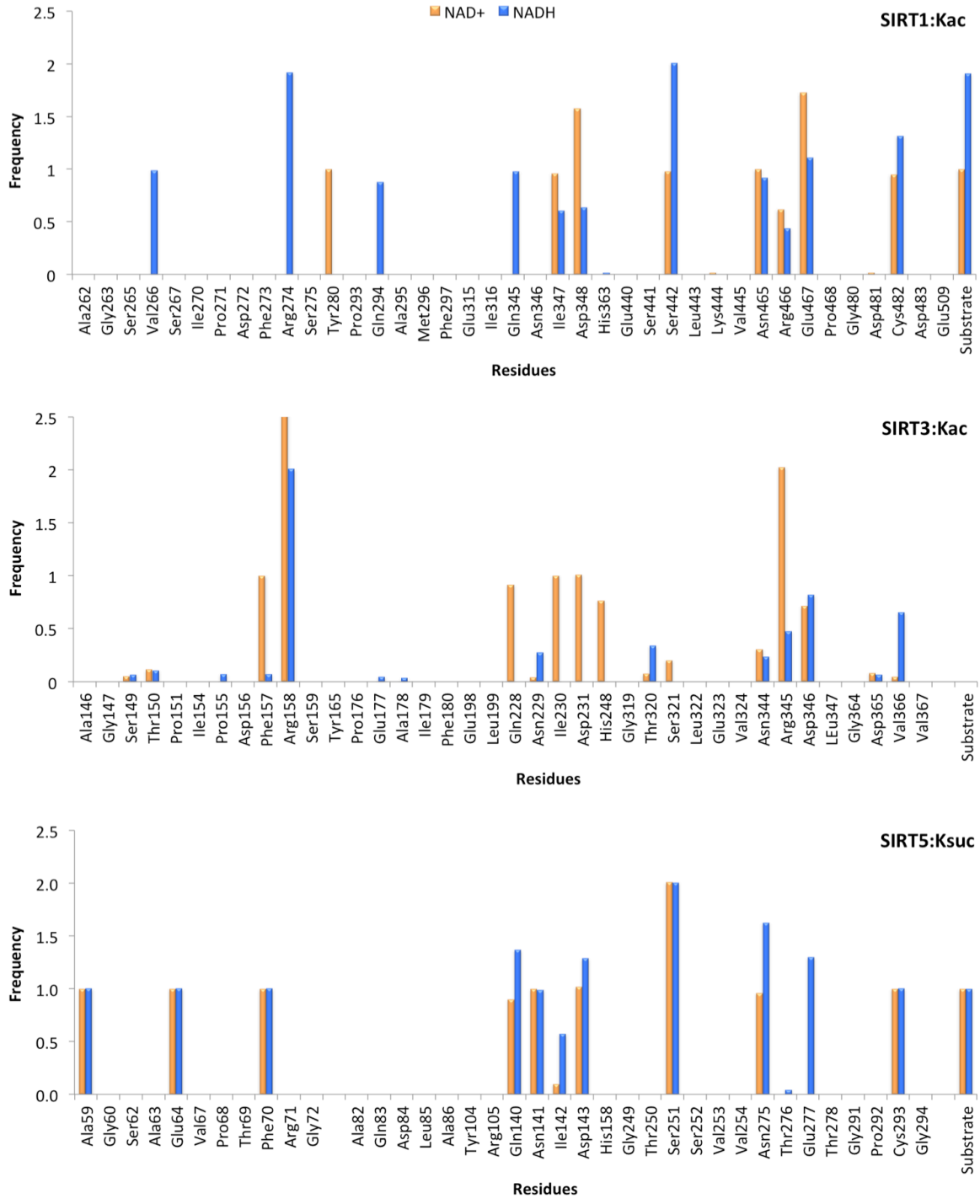


Figure 56 Hydrogen bonds between protein and dinucleotide, over the last 50ns of simulation time.

Positioning of the nicotinamide moiety in the C pocket and the binding to two highly conserved residues (I347/230/142 and D348/231/143 in SIRT1/3/5, hydrogen bonding

with the amide oxygen and proton, respectively) is essential for the deacylation. Both of these hydrogen bonds are frequently established between enzymes and NAD^+ but not NADH. In fact, these hydrogen bonds decrease when SIRT1 binds NADH and completely disappear in the case of SIRT3 due to the nicotinamide moiety movement. In the case of SIRT5, due to the very similar positioning to NAD^+, these interactions are also observed for NADH.

The differences between the NAD^+ and NADH binding were further investigated through the analysis of the nicotinamide moiety and the carboxamide group rotations during the simulations (Figure 57). In all three enzymes, NAD^+ adopts an *anti/syn* conformation that allows the binding to the conserved residues. However, although NADH can also adopt a productive *anti/syn* conformation in SIRT5, the *syn/syn* conformation is more frequent with SIRT1 and SIRT3.

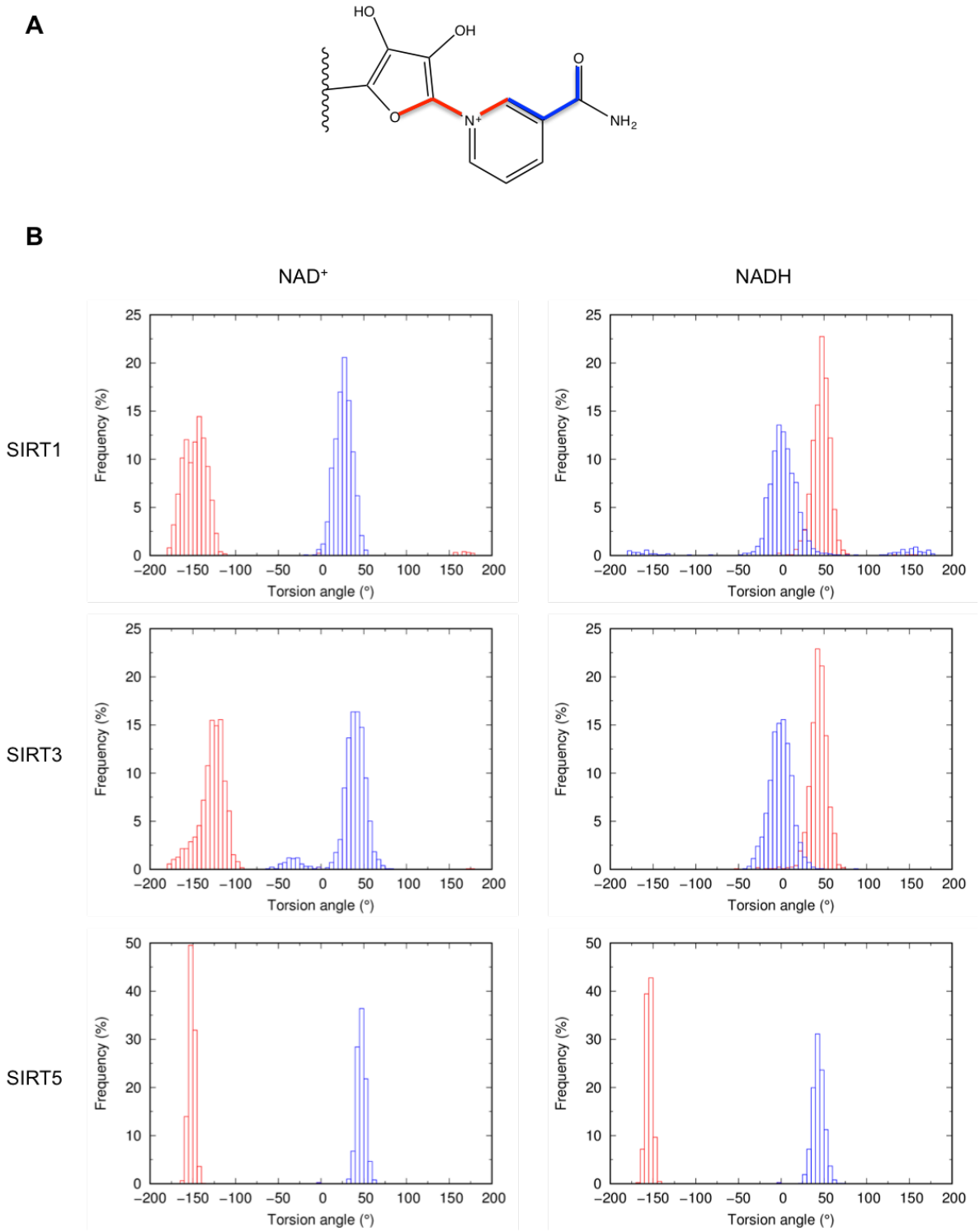


Figure 57 Analysis of nicotinamide and amide groups rotation. A. Schematic representation of the dihedral angles analysed in Figure 5. Colours correspond to the histograms. B. Histogram representation of the dinucleotide dihedral angles (as shown in Figure 4), during the last 50ns of simulation time.

In order to evaluate the likelihood of these conformations to occur, we ran a conformational search on the nicotinamide moiety, for both dihedral angles presented above. According to the contour plots in Figure 58, the NAD^+ moiety can find low energy conformations at any combination of *anti* and *syn* dihedral angles. On the other hand, the lowest energy conformation for the NADH moiety is indeed a *syn/syn* conformation.

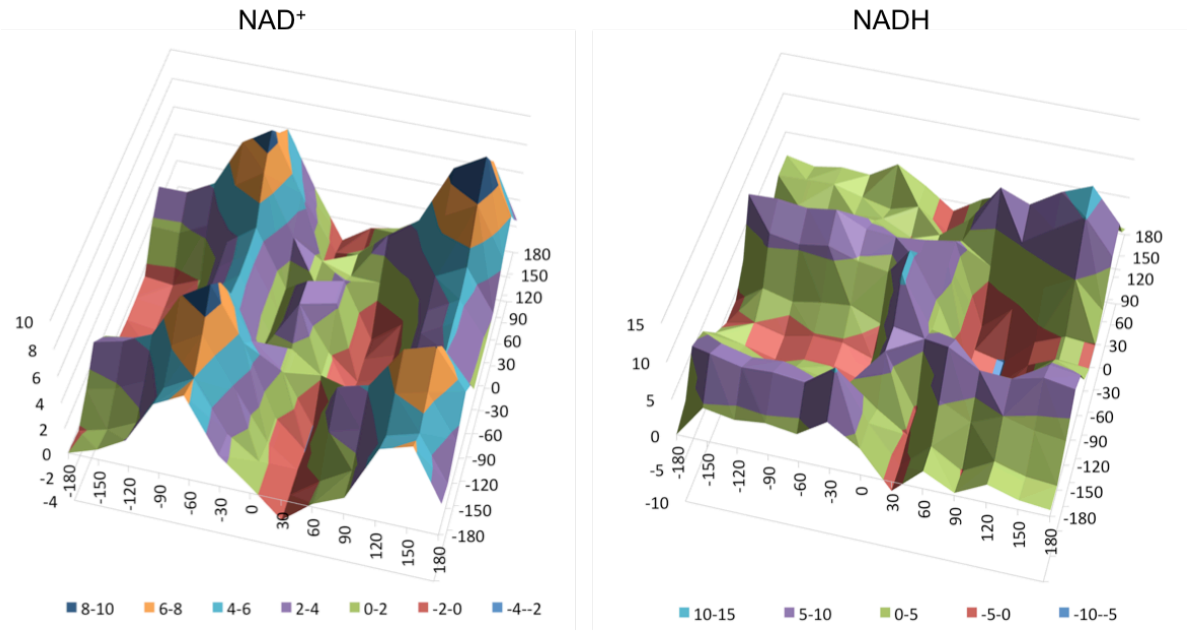


Figure 58 3D countour plots of the conformational search of the two dihedral angles represented in Figure 57A.

When in the *cis* conformation in relation to C1 in the ring, the amide is planar, which allows for a more favourable overlap between the orbitals of the nitrogen lone pair and the *pi*-system, resulting in a lower energy pose (Figure 59). In the *trans* conformation, the nitrogen is skewed such that an interaction between the lone pair and the hydrogen in the C1 position is established. The nitrogen rotation out-of-plane results in a higher energy conformation.

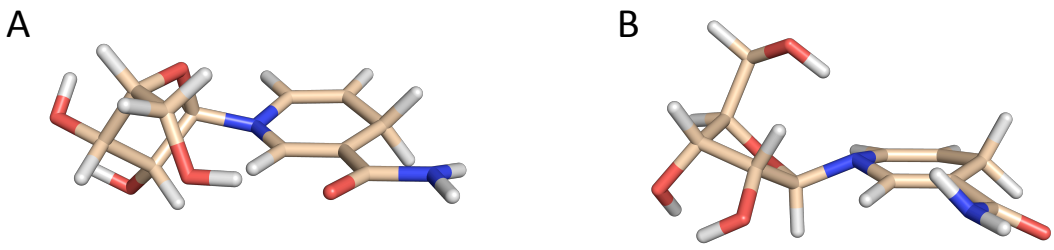


Figure 59 Structures from coordinate scan of NADH truncated model. Amide in cis (A) and trans (B) conformations in relation to C1.

3.1.2 AB pocket

In order to evaluate the stability of NAD+ and NADH in the unproductive conformation, the dinucleotides were placed in the AB pocket. For each one of the six systems (NAD+/NADH, SIRT1/SIRT3/SIRT5) I ran a 500ns MD simulation in Desmond, using the OPLS3 force field. The analysis of hydrogen bonds and torsion angles was performed over the last 50ns of simulation, when all the systems were equilibrated.

In agreement with the literature, the ADP-ribose moiety of both NAD+ and NADH remains in the A pocket of the two proteins during the simulation (Figure 60). In general, the nicotinamide moiety of the NADH tends to move further away from the B and C pockets, while the nicotinamide moiety of NAD+ tends to move closer to the C pocket, towards the productive binding conformation. In the case of SIRT1, this movement is evident and substantial, with NAD+ moving towards the C pocket and stabilizing between the two pockets (B and C). In SIRT3, there is a slight movement of NADH away from the B pocket and approaching the protein surface. On the other hand, NAD+ adopts a similar pose to the initial conformation and accommodates itself in the AB pocket. In SIRT5, the nicotinamide moiety of NAD+ also moves slightly towards the protein core in contrast to NADH which moves closer to the enzyme surface.